

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
CAMPUS ARARANGUÁ**

Gabriel Linardi

**UM MODELO DE DADOS ABERTOS APLICADO EM UM  
CENÁRIO GOVERNAMENTAL**

Araranguá, dezembro de 2013.

Gabriel Linardi

Gabriel Linardi

## **UM MODELO DE DADOS ABERTOS APLICADO EM UM CENÁRIO GOVERNAMENTAL**

Trabalho de Conclusão de Curso  
submetido à Universidade Federal de  
Santa Catarina, como parte dos  
requisitos necessários para a obtenção  
do Grau de Bacharel em Tecnologias  
da Informação e Comunicação.

Orientador: Prof. Dr. Alexandre  
Leopoldo Gonçalves.

Araranguá, dezembro de 2013.

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Linardi, Gabriel

UM MODELO DE DADOS ABERTOS APLICADO EM UM CENÁRIO  
GOVERNAMENTAL / Gabriel Linardi ; orientador, Alexandre  
Leopoldo Gonçalves - Florianópolis, SC, 2013.  
65 p.

Trabalho de Conclusão de Curso (graduação) -  
Universidade Federal de Santa Catarina, Campus Araranguá.  
Graduação em Tecnologias da Informação e Comunicação.

Inclui referências

1. Tecnologias da Informação e Comunicação. 2. Dados  
Abertos. 3. Dados Abertos Governamentais. 4. Resource  
Description Framework. 5. Tripla RDF. I. Gonçalves,  
Alexandre Leopoldo. II. Universidade Federal de Santa  
Catarina. Graduação em Tecnologias da Informação e  
Comunicação. III. Título.

Gabriel Linardi

## **UM MODELO DE DADOS ABERTOS APLICADO EM UM CENÁRIO GOVERNAMENTAL**

Este Trabalho de Conclusão de Curso foi julgado aprovado para a obtenção do Título de Bacharel em Tecnologias da Informação e Comunicação, e aprovado em sua forma final pelo Curso de Graduação em Tecnologias da Informação e Comunicação.

Araranguá, dezembro de 2013.

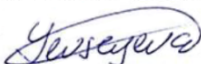


Prof. Vilson Gruber, Dr.  
Coordenador do Curso


### **Banca Examinadora:**



Prof. Alexandre Leopoldo Gonçalves, Dr  
Orientador  
Universidade Federal de Santa Catarina



Prof.ª Olga Yevseyeva, Dr.ª  
Universidade Federal de Santa Catarina



Prof. Juarez Bento da Silva, Dr.  
Universidade Federal de Santa Catarina

Este trabalho é dedicado a minha  
família e a todos que contribuíram  
para minha formação.

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus por me dar forças para não desistir nunca dos meus objetivos; a minha família que sempre me apoiou nas minhas escolhas; a todos que me apoiaram direta ou indiretamente; e em especial, agradeço ao Orientador Prof. Alexandre que, com muita paciência e profissionalismo me acompanhou durante esta fase possibilitando a realização deste trabalho.

## RESUMO

Nos últimos anos a disseminação da informação, pública ou privada, em formato padronizado vem ganhando importância. Isto gera demandas principalmente em cenários governamentais onde a transparência das ações públicas é cada vez mais exigida pela sociedade. Para tal, torna-se necessário o estabelecimento de modelos abertos de dados que permitam o compartilhamento e a integração com fontes de informação diversas. Entre estes modelos encontra-se o baseado em triplas, ou seja, Sujeito, Predicado e Objeto, sendo o RDF (*Resource Framework Description*) um dos mais utilizados. Além da disponibilização deve-se considerar a infraestrutura para o armazenamento e consulta dos dados, bem como, para possibilitar o desenvolvimento de aplicações variadas. Este trabalho propõe um modelo de dados em formato aberto capaz de representar uma fonte de informação governamental. A fonte escolhida foi o Site DivulgaCand disponibilizado pelo Tribunal Superior Eleitoral que possibilita a qualquer pessoa consultar os registros de candidatura da eleição de 2012. Tais informações estão disponíveis, mas não em formato recomendado para dados abertos. Para viabilizar a carga do modelo proposto foi desenvolvido um protótipo que de maneira genérica efetua a leitura de uma fonte de dados origem e transfere as informações para uma base de dados triplificada. O desenvolvimento do trabalho permite assim afirmar que, apesar da iniciativa de dados abertos não ser algo tão recente, faltam exemplos principalmente no meio privado. Conclui-se também que existe, a partir dos dados abertos, potencial para o desenvolvimento de serviços com impacto na sociedade. Mais especificamente em relação ao modelo proposto, o mesmo se mostra adequado e capaz de responder a perguntas em linguagem apropriada permitindo assim um melhor entendimento do domínio de registro de candidaturas.

**Palavras-chave:** Dados Abertos; Dados Abertos Governamentais; Resource Description Framework, Tripla RDF.

## ABSTRACT

In recent years the dissemination of information, public or private, in a standard format has gained relevance. It creates demands in scenarios where government transparency of public actions is increasingly required by society. Thus, it is necessary to establish open data models that enable sharing and integration of diverse information sources. Among these models we can find those based on triples, that is, Subject, Predicate and Object, in which the RDF (Resource Description Framework) is one of the most used. Beyond the availability of the data we should consider the infrastructure for the storage and retrieval of data and to enable the development of several applications. This work proposes a model of open data capable of representing a government information source. The information source chosen was the site *DivulgaCand* provided by the *Tribunal Superior Eleitoral* that allows anyone to query the 2012 election data. Such information is available but not in a recommended format to open data. Aiming to load the proposed model it was developed a prototype which generically read data from a data source and transfers it to a triple store. The development of the work enables to state, although the open data initiative is not so recent, that examples in the private environment are sparse. We also conclude that there is, from the open data perspective, potential for the development of services with impact on society. Taking into account the proposed model, it proves to be suitable and able to answer questions in an appropriate language thus allowing a better understanding of the application registration domain.

**Keywords:** Open Data; Open Government Data; Resource Description Framework; RDF Triple.



## LISTA DE FIGURAS

Figura 1- Exemplo de grafo de classificação de carros.....	30
Figura 2 - URIs usadas para descrever pessoas e as relações entre elas. ....	31
Figura 3 - Exemplo de endereço de URL.....	32
Figura 4 - Arquitetura Jena. ....	33
Figura 5 - Ciclo de disponibilização de dados abertos.....	42
Figura 6 - Pagina principal do Site DivulgaCand. ....	43
Figura 7 - Informações de candidatos disponibilizadas por estado e município. .....	44
Figura 8 - Modelo proposto para a disponibilização dos dados do Site DivulgaCand.....	45
Figura 9 - Modelo Entidade-Relacionamento proposto para suportar os dados originais do Site DivulgaCand. ....	46
Figura 10 - Exemplo de grafo para a informação de um candidato em particular. .....	48
Figura 11 - Exemplo de grafo para Cidade. ....	49
Figura 12 - Protótipo desenvolvido para possibilitar a migração dos dados. ....	50

## LISTA DE QUADROS

Quadro 1 - Exemplo RDF sobre classificação de carros. ....	30
Quadro 2 - Método para criação de modelo.....	34
Quadro 3 - Cláusulas básicas da linguagem SPARQL. ....	35
Quadro 4 - Consulta SPARQL básica.....	36
Quadro 5 - Exemplo de uso de Prefixo. ....	36
Quadro 6 - Clausula OPTIONAL. ....	37
Quadro 7 - Clausula UNION. ....	37
Quadro 8 - Cláusulas ORDER BY, LIMIT e OFFSET. ....	38
Quadro 9 - Cláusula UNBOUND. ....	38
Quadro 10 - A clausula de negação. ....	38
Quadro 11 - Cláusula ASK. ....	39
Quadro 12 - Exemplo em RDF/XML de um candidato e demais informações relacionadas. ....	47
Quadro 13 - Método utilizado para estabelecer as conexões com os bancos de dados de origem e destino. ....	51
Quadro 14 - Método utilizado para finalizar as conexões com os bancos de dados de origem e destino. ....	51
Quadro 15 - Método para criação de estrutura no banco de Dados. ....	52
Quadro 16 - Procedimento para localizar a meta informação no banco de dados de origem (Relacional). ....	53
Quadro 17 - Função para realizar a Inserção no banco de dados RDF. ....	54
Quadro 18 - Seleção de algumas cidades da base de triplas. ....	55
Quadro 19 - Resultado da seleção de Cidades. ....	55
Quadro 20 - Seleção de informações de alguns estados. ....	55
Quadro 21 - Resultado da seleção das informações de estados. ....	56
Quadro 22 - Seleção das informações do candidato 25777. ....	56
Quadro 23 - Resultado da seleção do candidato 25777. ....	56
Quadro 24 - Seleção considerando informações de candidatos e cidades. ....	57
Quadro 25 - Resultado da seleção envolvendo candidatos e cidades. ....	57
Quadro 26 - Seleção utilizando os prefixos de cidades e estados. ....	58
Quadro 27 - Resultado da seleção envolvendo cidades e estados.....	58

## **LISTA DE ABRACIATURAS E SIGLAS**

API - Application Programming Interface.  
CERN - European Organization for Nuclear Research.  
CSV - Comma-separated values.  
HTML - HyperText Markup Language.  
JSON - JavaScript Object Notation.  
NCE - Núcleo de Computação Eletrônica.  
OD4D - Open date for development.  
OWL - Web Ontology Language.  
PICS - Platform for Internet Content Selection.  
RDF - Resource Description Framework.  
RDFS - Resource Description Framework Schema.  
SPARQL - SPARQL Protocol and RDF Query Language.  
SWRL - Semantic Web Rule Language.  
URI - Uniform Resource Identifier.  
URL - Uniform Resource Locator.  
W3C - World Wide Web Consortium.  
XML - eXtensible Markup Language.

# SUMÁRIO

1 INTRODUÇÃO.....	13
1.1 JUSTIFICATIVA .....	16
1.2 OBJETIVOS .....	17
1.2.1 Objetivo Geral .....	17
1.2.2 Objetivos Específicos .....	18
1.3 METODOLOGIA.....	18
1.4 ESTRUTURA DO TRABALHO .....	19
2 DADOS ABERTOS .....	20
2.1 WEB SEMÂNTICA .....	22
2.1.1 Metadados .....	24
2.2 INFRAESTRUTURA.....	25
2.2.1 RDF .....	25
2.2.1.1 A origem do RDF .....	27
2.2.1.2 Modelo de dados RDF .....	28
2.2.1.3 Estrutura do RDF.....	29
2.2.1.4 Namespace.....	31
2.2.2 Jena.....	32
2.2.2.1 Criação de modelos .....	33
2.2.3 SPARQL.....	34
2.2.3.1 Exemplos de consultas em SPARQL.....	35
2.3 POSSIBILIDADES DE DADOS ABERTOS .....	39
3 CENÁRIO DE APLICAÇÃO E MODELO PROPOSTO .....	42
3.1 PORTAL DO CANDIDATO .....	43
3.2 MODELO PROPOSTO.....	44
3.2.1 Modelo Relacional.....	45
3.2.2 Ferramenta de Conversão .....	46
3.2.3 Modelo RDF.....	46
4 DESENVOLVIMENTO DO PROTÓTIPO E AVALIAÇÃO DOS RESULTADOS .....	50
4.1 DETALHAMENTO DO PROTÓTIPO.....	50
4.2 AVALIAÇÃO DOS DADOS DISPONIBILIZADOS .....	54
5 CONSIDERAÇÕES FINAIS .....	59
REFERÊNCIAS .....	62

## 1 INTRODUÇÃO

A internet atual pode ser definida como a Web Sintática. Nesta rede os computadores fazem apenas a apresentação das informações, porém o processo de interpretação é de responsabilidade de cada pessoa que deseja utilizar esses conteúdos. O processo de interpretação é algo muito mais difícil e trabalhoso e requer um grande esforço para avaliar, classificar e selecionar as informações de que se tem interesse. Sendo assim, chega-se a seguinte questão: Por que os computadores não podem executar serviços de interpretação?

A razão da dificuldade de interpretação de informações pelas máquinas é por não possuírem informações suficientes que as descreva semanticamente de modo que interpretações possam ser realizadas. Como exemplo pode-se citar as consultas realizadas atualmente através de mecanismos de busca tradicional, ao se pesquisar por “Jaguar” o usuário recebe a página de maior relevância, mas este resultado não considera o real desejo do usuário, pois “Jaguar” pode ser um animal ou um carro.

Deste modo, no atual estágio de tecnologia os computadores têm a função de direcionamento e entrega de informações, não sendo capazes de analisar de maneira mais ampla o conteúdo das páginas, uma vez que as informações estão estruturadas para serem utilizadas por pessoas e não por máquinas. Este é um dos motivos pela qual se enfrenta dificuldades quando se realiza uma consulta sobre determinado assunto e o que realmente se procura não é encontrado de modo fácil e da maneira desejada. Segundo Berners-Lee, Hendler e Lassila (2001), o modelo atual é chamado de Web de documentos e não de informações, por isso, os computadores oferecem ajuda limitada no acesso e processamento da informação, deixando as funções mais complexas, como a interpretação dessa informação a cargo dos seres humanos.

Com estas premissas vem se desenvolvendo nos últimos anos o conceito de Web Semântica visto como uma evolução da Web atual ou a Web de Documentos. Esta nova proposição poderá oferecer estrutura capaz de permitir que dados e informações sejam compartilhadas e reutilizadas por aplicações computacionais de maneira mais inteligente.

A Web Semântica é um projeto liderado pela W3C (World Wide Web Consortium) e tem envolvido um grande esforço de grupos de pesquisas de diferentes áreas (FENSEL et al., 2005). Além disso, a Web Semântica busca unir vários domínios para permitir que se alcance um entendimento compartilhado de diversos contextos. A Web Semântica visa permitir que os dados e conteúdos da Web sejam interpretados tanto por humanos quanto por sistemas informatizados, permitindo que programas de computadores tenham maior capacidade de interpretação de conteúdos visando a realização de tarefas

complexas. Ela estende a Web Clássica ou Web de Documentos através da inserção de metadados nas páginas promovendo assim meios para se tornarem legíveis para máquinas.

O modelo atual utilizado da Web foi proposto por Tim Berners-Lee quando ele trabalhava para o CERN (Organizações Européia para a Pesquisa Nuclear), em 1989, com a ideia de compartilhar os documentos de pesquisa entre seus colegas de trabalho (W3C, 2008b). Durante este período, Berners-Lee identificou algumas deficiências no modelo atual de Web e em 2001 publicou um segundo artigo juntamente com outros autores em que foi proposta uma nova forma de se utilizar a Web, uma forma mais próxima de utilizá-la em sua total potencialidade (BERNERS-LEE; HENDLER; LASSILA, 2001).

Deste modo, foram iniciadas novas pesquisas para que se fosse criado um formato de Web onde os dados estariam mais organizados. Soma-se a isto o fato dos mecanismos de busca enfrentarem dificuldades para realizar pesquisas mais precisas entre documentos, uma vez que a diferenciação em termos de assunto, qualidade e relevância não é uma tarefa fácil. A tecnologia atual oferece pouca capacidade de diferenciar uma informação comercial de uma educacional ou informação entre idiomas e culturas. É necessário haver informações da própria informação, esta informação que é chamada de metadados possui o objetivo de classificar a informação e tornar os processos de busca mais eficazes.

Portanto, algumas dessas novas estruturas necessárias já foram definidas e outras ainda estão sendo desenvolvidas pelo W3C. A W3C constituída por um consórcio de organizações interessadas na definição e desenvolvimento de novos conceitos, protocolos e padrões de estruturas para a Web, onde visam obter maior eficácia de seus recursos, a Web Semântica utiliza de padrões e linguagens tais como, XML (*eXtensible Markup Language*), XML Schema, RDF (*Resource Description Framework*), RDF Schema, OWL (*Web Ontology Language*), SWRL (*Semantic Web Rule Language*) entre outras, estes modelos de representação de informação utilizados para aumentar a usabilidade e funcionalidade da web e de seus recursos relacionados (SHADBOLT; HALL; BERNERS-LEE, 2007).

Para que fosse possível demonstrar e trabalhar com as informações dispostas na Web de uma maneira mais estruturada, criou-se o RDF. RDF é uma linguagem para representar informações na Internet recomendada pela W3C que tem como objetivo simplificar o modelo de dados, utilizando o conceito de URI (*Uniform Resource Identifier*) (W3C, 2008a).

A adoção de novos padrões e tecnologias é necessária para possibilitar o conceito de Web Semântica com o intuito de aumentar a usabilidade e funcionalidades, não somente do conteúdo atualmente disponível na Web, mas também dos dados mantidos por organizações

públicas ou privadas. Com a utilização de padrões atualmente em desenvolvimento é possível que as informações sejam disponibilizadas em formatos úteis permitindo que qualquer pessoa ou organização possa utilizá-las, promovendo assim o conceito de dados abertos.

Este conceito não é uma novidade, mas cria um contexto para que qualquer pessoa interessada possa utilizar determinada informação sem restrição visando facilitar o entendimento e a análise de informações (GRAY et al., 2012).

Os dados disponibilizados de forma aberta são vistos como uma nova maneira de compartilhar dados e documentos entre pessoas e entidades na Web com finalidades variadas. Para a criação de aplicações e serviços não basta apenas ter dados disponíveis livremente, é preciso que esses dados sejam úteis, e que possam ser processados por máquinas. Ou seja, os dados precisam ser disponibilizados em formatos adequados para as aplicações que irão se utilizar dos mesmos. E isso é importante tanto para produtores quanto para consumidores de informação (BERNERS-LEE; HENDLER; LASSILA, 2001).

A utilização de dados abertos busca estabelecer novos planos para a disponibilização de conteúdo coletivo que possibilite uma maior interoperabilidade entre sistemas. Com a disponibilização de dados abertos novos produtos e serviços podem ser criados por organizações públicas ou privadas, e mesmo por cidadãos comuns. De modo geral a disponibilização de dados abertos tem como finalidade o fato de que qualquer organização ou pessoa pode se utilizar, distribuir e/ou criar aplicações com base nestes dados sem qualquer restrição (AUER et al., 2007). O termo “dado aberto” (do inglês *open data*) ganhou popularidade em 2009 no evento conhecido como TED® onde Tim Bernes-Lee fez um apelo “Dado Aberto já”, solicitando que empresas, governos e cientistas disponibilizassem seus dados de forma gratuita.

No nível tecnológico existem vários padrões e formatos que permitem a disponibilização em formato aberto. No entanto, Tim Bernes-Lee defende o uso de padrões baseados na Web Semântica (BERNERS-LEE, 2006), em que os dados estão conectados (*linked*) e possuem informações adicionais visando permitir que humanos e máquinas explorem os conteúdos disponíveis na Web. Para disponibilizar estes dados, empresas, governos e instituições deverão enfrentar grandes mudanças no modo de trabalho. Além disto, devem ter a consciência de que tudo que for disponibilizado será visível e utilizado por todos. Entre as instituições que vem disponibilizando seus dados encontram-se as governamentais, com o intuito de tornar públicas as informações sobre os gastos, recursos e investimentos realizados, promovendo assim uma maior transparência de seus atos.

Segundo Diniz (2010), os dados abertos podem ainda ajudar na inclusão digital pelo motivo de fornecerem um meio padronizado e

acessível permitindo a visualização do conteúdo com ferramenta especializada por qualquer pessoa.

A disponibilização dos dados abertos governamentais possui um papel importante na responsabilidade de cada governo, pois o acesso facilitado aos dados do governo habilita as pessoas a monitorarem os atos dos governos criando assim, um vínculo entre cidadão e governo para atingir metas em suas políticas públicas. Um dos desafios para o crescimento da democracia é o desenvolvimento contínuo de novas ferramentas para unir cada vez mais o governo e sociedade, com as novas tecnologias da informação e comunicação, tendo ao seu lado a Internet como ponto central de suporte os governos podem estabelecer condições ideais para o acesso e a utilização dos dados públicos para os cidadãos.

Para que estes dados sejam disponibilizados na Web de maneira correta, de modo que os usuários e os computadores possam entender por completo o conteúdo disponibilizado é necessária a aplicação de novas regras. Para isso, Berners-Lee sugere novas práticas para publicação de dados na Web com o objetivo de reduzir a diversidade atualmente encontrada (BERNERS-LEE, 2006).

Segundo Berners-Lee (2006), para que se inicie este novo padrão de disponibilização dos dados alguns critérios devem ser seguidos, entre eles: (a) a disponibilização dos dados em seu formato mais bruto possível, ou seja, antes de qualquer agregação; (b) a disponibilização em formato aberto não proprietário, estável e de amplo uso; e (c) a não existência de instrumento jurídico que impeça sua reutilização. A adoção destas práticas e padrões objetiva tornar a Web um espaço interconectado mais rico em sua descrição promovendo assim uma maior utilidade e aplicabilidade dos dados, e permitindo que determinado conteúdo de interesse seja mais facilmente identificado.

## 1.1 JUSTIFICATIVA

Dentre as características de dados abertos, menciona-se que estes dados precisam e devem ser úteis e passíveis de serem processados por máquinas, isto é importante tanto para quem está do lado da produção destes dados, quanto para as pessoas que irão consumi-las.

Existem muitos padrões e formatos que permitem esse tipo de disponibilização. Os modelos mais comuns são CSV, XML, JSON, APIs, DBs etc. No entanto, Tim Berners-Lee (2006), o criador da Web, defende o uso de um novo modelo que será utilizado para a publicação de dados, o *RDF*. O modelo RDF é vista como uma linguagem utilizada para realizar a representação de informações na Web, sendo uma especificação para representar documentos e metadados (W3C, 2004). O principal objetivo do RDF é apresentar uma codificação junto com um



mecanismo de interpretação, para que softwares específicos sejam capazes de entender o conteúdo de páginas Web (ANTONIOU; HARMELEN, 2008).

A partir destes padrões a abertura de dados, seja por instituições públicas ou privadas, torna-se mais útil, permitindo uma maior democratização da informação. Entre as instituições vem se destacando os governos de vários países ao redor do mundo, incluindo o governo brasileiro.

O governo brasileiro esta apostando neste modo de disponibilização de dados nos ambientes Web para que estes possam ser utilizados publicamente pelo cidadão comum ou outras organizações. Informações públicas compartilhadas em formato bruto e abertas poderão ser reutilizadas e transformadas para atender as necessidades do mais diversos tipos e diferentes possibilidades de análise da informação.

O Brasil é pioneiro quando se fala em dados abertos na América Latina. Para manter esta posição de referência, o país deverá se focar em três importantes passos: a) publicar informações governamentais na Web; b) construir a cultura de governo aberto; e c) criar políticas favoráveis para guiar e consolidar este processo (MAZONI, 2011). Ainda segundo Mazoni (2011), o país só tem a ganhar com a disponibilização de dados dos governos para a sociedade.

Independente do domínio, financeiro, demográfico, orçamentário, educacional entre outros, desde que estes dados sejam acessíveis em formato aberto na Web, isso potencializa a democratização da informação e a geração de novos conhecimentos, soma-se a isto o aumento da participação civil nas decisões políticas, e deste modo, se aperfeiçoa os serviços prestados e se estimula o crescimento da sociedade de determinado país, uma vez que seus cidadãos passam a ter mais consciência dos fatos (MAZONI, 2011).

Partindo destes pressupostos a disponibilização de modelos e aplicações que auxiliem no processo de geração de dados abertos a partir de fontes de informações governamentais se tornam relevantes, e são justificáveis principalmente pela democratização e possibilidade de desenvolvimento de aplicações diversas.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Propor um modelo voltado a dados abertos a partir de uma fonte de informação governamental com o intuito de democratizar a informação e possibilitar a interoperabilidade entre sistemas computacionais.

### 1.2.2 Objetivos Específicos

Visando atingir o objetivo principal, alguns objetivos específicos são requeridos, entre eles:

- Identificar na literatura o ferramental necessário para estruturar e persistir informações de fontes de informações em formato de dados abertos, para posterior recuperação e apresentação;
- Estudar um domínio de aplicação em que seja possível a proposição de um modelo de dados abertos;
- Propor um modelo de dados que permita representar o domínio de aplicação estudado;
- Desenvolver um protótipo que possibilite a realização da tarefa de conversão dos dados de um domínio de aplicação para um modelo de dados aberto;
- Avaliar os dados disponibilizados em formato aberto através de um conjunto de consultas.

### 1.3 METODOLOGIA

O presente trabalho é de base tecnológica e envolve a proposição de um modelo de publicação de dados abertos a partir de dados retirados do Site Divulgacand<sup>1</sup> referente ao ano de 2012. A metodologia empregada neste trabalho é dividida em cinco etapas.

Etapa 1: Será realizada a pesquisa do referencial teórico abordado no trabalho envolvendo modelos, linguagens e ferramentas para a representação e manipulação de dados abertos;

Etapa 2: Pesquisa de material teórico sobre o domínio de aplicação deste trabalho;

Etapa 3: Criação de uma base de dados relacional baseado no Site Divulgacand, detalhando os componentes e informações mais importantes disponibilizadas;

Etapa 4: Proposição do modelo de dados abertos em formato padronizado capaz de representar o banco relacional criado para armazenar os dados do Site DivulgaCand.

Etapa 5: Desenvolvimento de um protótipo para conversão da base de dados relacional em uma base de dados no modelo utilizado por dados abertos, ou seja, uma base de dados em triplas.

Etapa 6: Elaboração e aplicação de consultas sobre a base de dados em triplas de modo que se possa avaliar o modelo proposto.

---

<sup>1</sup> <http://divulgacand2012.tse.jus.br/divulgacand2012>.

## 1.4 ESTRUTURA DO TRABALHO

O documento está dividido em 4 Capítulos, além da introdução, sendo os demais Capítulos relacionados abaixo.

O Segundo Capítulo é responsável por apresentar de forma geral os conceitos sobre dados abertos, as ferramentas e linguagens necessárias para que seja possível a publicação destes dados de forma correta. São mencionadas ainda algumas organizações que fazem uso do modelo de publicação de dados na Web.

No Terceiro Capítulo é apresentado o cenário de aplicação de que trata este trabalho, o site Divulgacand, bem como, o modelo proposto. O Quarto Capítulo detalha o protótipo desenvolvido para auxiliar na criação do modelo e apresenta algumas consultas elaboradas visando permitir uma avaliação do conteúdo disponibilizado.

Por fim, o Quinto Capítulo contém as considerações finais e os trabalhos futuros.

## 2 DADOS ABERTOS

O conceito de dados abertos vem sendo utilizado nos mais diferentes cenários em que exista a necessidade de disponibilização da informação para qualquer pessoa sem restrição de uso desta, ou seja, qualquer pessoa pode ter acesso a estes dados e documentos sem discriminação de grupos ou pessoas. Estes dados devem ser disponibilizados em formatos padronizados para que o entendimento e a análise das informações sejam facilitados (GRAY et al., 2012).

Através dos dados abertos são esperadas diversas vantagens no processo de disponibilização e utilização desses dados (ZUIDERWIJK; JANSSEN; PARNIA, 2013). Os dados abertos estão sendo universalmente reconhecidos, pois desenvolvedores vem elaborando projetos onde o objetivo é criar um “Mundo” de informações estruturadas, para que seja possível realizar consultas semanticamente ricas (AUER et al., 2007).

O entendimento sobre este formato estruturado relaciona-se a informação pública ou não, que qualquer cidadão possui acesso e permissão para reutilização em qualquer propósito. Em 2009, os dados abertos começaram a se tornar visíveis ao público em geral na medida em que governos de vários países, entre eles, EUA, Reino Unido, Canadá e Nova Zelândia, anunciaram novas iniciativas de abertura de suas informações públicas (GRAY et al., 2012).

Atualmente, quando se discute o acesso à democracia e as políticas públicas, fala-se sobre transparência de informações e ao acesso a documentos são produzidas pelos Governos. Claramente, o uso da Internet para acesso a estas informações tem sido considerado como o meio ideal pelas pessoas que tem interesse em acompanhar as informações sobre o governo, e a ele se convencionou chamar de governo eletrônico (MAZONI, 2011).

Contudo, este cenário promove grandes dificuldades levando-se em conta que a sociedade demanda cada vez mais transparência dos atos dos governos. Por este motivo, o simples acesso às informações já não é suficiente, pois a capacidade de compartilhar e reutilizar estas informações são fator determinante para acompanhar as ações de governo.

Para isso se estuda o uso de dados abertos que são os dados e as informações disponibilizadas na Web, compartilhados em formato bruto e aberto, compreensíveis logicamente, de modo a permitir sua reutilização em aplicações digitais desenvolvidas pela sociedade (W3C, 2011). As informações e dados do governo que são disponíveis para a sociedade na Web, consiste na publicação de dados sobre as informações públicas em formatos que permitam o seu compartilhamento, acesso, descoberta e fácil manipulação pelos consumidores desses dados (BREITMAN, 2006).

Segundo Eaves (2009), dados abertos governamentais são regidos por três leis básicas:

- Se o dado não pode ser encontrado e indexado na Web, ele não existe;
- Se não estiver aberto e disponível em formato compreensível por máquina, ele não pode ser reaproveitado;
- Se algum dispositivo legal não permitir sua replicação, o dado não é útil.

Além das três leis dos dados abertos governamentais, existem ainda oito princípios básicos. Segundo a “*Open Government Working Group*” (OPENGOVDATA, 2007), para que os dados no contexto governamental sejam considerados abertos estes deve ser:

1. Completos: Todos os dados públicos estão disponíveis. Dado público é o dado que não estará sujeito a limitações válidas de privacidade, segurança ou controle de acesso;
2. Primários: Os dados são apresentados como os coletados na fonte, com o maior nível possível de granularidade e sem agregação ou modificação;
3. Atuais: Os dados são disponibilizados tão rapidamente quanto necessária visando a preservação do seu valor;
4. Acessíveis: Os dados são disponibilizados objetivando o maior alcance possível de usuários e para o maior conjunto possível de finalidades;
5. Compreensíveis por máquinas: Os dados são razoavelmente estruturados de modo a possibilitar processamento automatizado;
6. Não discriminatórios: Os dados são disponibilizados para todos, sem exigência de requerimento ou cadastro;
7. Não proprietários: Os dados são disponíveis em formato sobre o qual nenhuma entidade detenha controle exclusivo;
8. Livres de licenças. Os dados não estão sujeitos a nenhuma restrição de direito autoral, patente, propriedade intelectual ou segredo industrial. Restrições sensatas relacionadas à privacidade, segurança e privilégios de acesso são permitidas.

Para por em prática esta proposta de dados governamentais abertos, uma das alternativas tecnológicas está baseada no uso da Web Semântica utilizando todas as tecnologias associadas a ela, incluindo as tecnologias que são sugeridas pelo Memorando de Transparência e Dados Abertos (no inglês *Memorandum on*

*Transparency and Open Government*) que afirma que a abertura destes dados vão fortalecer a democracia e promover a eficiência e efetividade do governo (HOLDREN; ORSZAG; PROUTY, 2009).

## 2.1 WEB SEMÂNTICA

A Web Semântica é vista como a “Web de dados” permitindo que máquinas possam entender a semântica, ou o significado, das informações na Web (*World Wide Web*). A Web Semântica estende a Web clássica (ou Web de documentos), seria a rede de *hyperlinks* de páginas Web, que é legível para humanos, inserindo informações “*metadados*” nestas páginas, que são legíveis por máquinas, onde contém informações sobre as páginas e como elas se relacionam umas com as outras, permitindo que os agentes automatizados acessem as informações da Web de forma mais inteligente. O papel da Web Semântica é organizar os dados da web, como o HTML fez para os sistemas de informação textuais, ou seja, prover flexibilidade suficiente para permitir representar os conteúdos das páginas de maneira que ferramentas consigam interpretar estas informações, e regras lógicas. (BERNERS-LEE, 2006)

A proposta da Web Semântica elaborada por Berners-lee partiu do princípio de que computadores não conseguem interpretar todo o significado do conteúdo de páginas Web. Ainda que existam sistemas computacionais capazes de realizar processamento especializado imitando o comportamento humano, ainda existem muitas informações na web que ainda são interpretadas somente por humanos, isto impossibilita que computadores identifiquem semanticamente os documentos, textos, páginas, imagens e demais conteúdos disponíveis na Web (BERNERS-LEE; HENDLER; LASSILA, 2001).

Um exemplo das implicações disto está nos resultados de buscas, muitas vezes distantes dos resultados esperados, já que um termo, palavra ou expressão pode ter diversos significados, dependendo do contexto. Isso ocorre pelo motivo dos dados estarem publicados em um formato incapaz de ser interpretado por máquinas, acreditava-se, que será possível criar uma estrutura capaz de adicionar semântica aos conteúdos da WEB (BREITMAN, 2006).

Criando páginas com informações e dados semanticamente corretas, seria possível o reconhecimento e a interpretação de contexto e significado por parte de sistemas computacionais, podendo ser usado para melhorar e enriquecer as buscas por conteúdo por estes sistemas inteligentes, possibilitando resultados mais relevantes, permitindo uma maior interação entre sistemas de informação e aumentando a interoperabilidade entre os mais diversos sistemas.

Por motivo destas informações não estarem em um formato adequado, e estarem armazenados em locais isolados, muitos

armazenados em bases proprietárias e em formatos fechados os governos de todo o mundo, incluindo o Brasil, estão enfrentando problemas para realizar a interação entre estes dados, pois com o passar do tempo o volume de dados cresceu muito dificultando ainda mais a interoperabilidade dos dados entre diferentes departamentos governamentais. Isto tem conduzido muitos governos a investirem em iniciativas de interoperabilidade baseada em padrões abertos (BATISTA; SILVA; MIRANDA, 2012).

Segundo Souza e Alvarenga (2004), a interoperabilidade é entendida como a possibilidade de sistemas de diferentes fontes se comunicarem. Os conteúdos publicados na Internet, só poderão ser integrados e reutilizados quando estes estiverem ancorados por algum tipo de sistema que possa permitir um alto grau de interoperabilidade (SAYÃO; MARCONDES, 2008).

Por este motivo, visando um novo modo de organização de dados na Web que pesquisadores de diversas áreas, mas principalmente da área de Inteligência Artificial, vem propondo uma série de modelos para disponibilização destes dados, a ideia principal é disponibilizar as informações de maneira padronizada com o objetivo de facilitar o acesso a informações e conteúdos na Web. A solução é similar à utilizada para classificar os seres vivos, em que biólogos se utilizam de taxonomias, estruturas de conhecimento de determinado domínio muito bem definidas, que são adotadas e compartilhadas por muitos pesquisadores ao redor do mundo. Desta forma, existem muitos esforços do sentido de criar um modelo estruturado para que se possa ter uma organização das informações e conteúdos na Web (BREITMAN, 2006).

Pode-se perceber que existem vários tipos de páginas das mais sofisticadas até as mais simples, criadas por pessoas leigas, que as utilizam para postagens pessoais e propagandas de seus negócios, disponibilização de informações familiares e até mesmo de seus bichos de estimação. Por este motivo é difícil imaginar que se possa ter um único modelo para organizar os dados na Web, uma vez que as pessoas que utilizam não possuem os mesmos conhecimentos para navegação e uso destas informações (BREITMAN, 2006).

Segundo Berners-Lee, Hendler e Lassila (2001), no futuro, ao invés da adoção de um modelo único de organização de informação, haverá uma série de modelos de organização de forma paralela. Desta maneira, qualquer empresa, organização ou universidade participante na chamada Web do futuro vai possuir seu próprio modelo de organização de dados e informações.

Uma das áreas que poderá se utilizar deste padrão é o Comércio Eletrônico, onde a consulta de preços a diversos fornecedores é bastante complexa. Atualmente os interessados devem fazer a comparação dos preços visitando todos os sites manualmente o que

torna uma compra online muito demorada. Nem sempre as condições, preços e prazos de entrega são demonstradas de maneira consistente e atualizada. Por exemplo, algumas empresas cobram seus fretes por peso e outra pela quantidade comprada.

Para esta classe de aplicação baseada na comparação de preços, existem os robôs chamados de *shopbots*, estes robôs realizam esta tarefa através de técnicas onde se objetiva ler todas as informações possíveis referentes a determinado produto, tais como, a descrição do produto, o preço e valor do frete. Contudo, como as informações das páginas não possuem um formato padronizado, muitas vezes estes robôs somente conseguem obter parte da informação, por exemplo, o nome do produto e seu preço. No caso do valor do frete esta informação somente poderá ser obtida após alguns passos do processo de compra. Sendo assim, a implementação destes robôs é complicada, pois ajustes e reprogramações são necessários cada vez que uma loja muda seu *layout* de página.

A comparação de preços e condições de venda de maneira mais confiável somente será possível quando todas as lojas disponibilizarem seus catálogos de produtos de modo processável pelos computadores, ou seja, através de estruturas padronizadas que explicitem o conhecimento de domínio utilizado no processo de negociação (BREITMAN, 2006).

### **2.1.1 Metadados**

Metadados são dados sobre os dados, ou seja, são os dados que propiciam informações adicionais para que computadores saibam de que assunto as páginas e conteúdos tratam.

Segundo Breitman (2006), a organização do conhecimento é um problema muito antigo, que já preocupava os filósofos da antiguidade, onde foi proposto o primeiro modelo de classificação utilizado para organizar as espécies em tipos e subtipos. A utilização de metadados na web traz um novo problema que antes não existia, pois quase todos os sistemas que utilizavam os metadados estavam limitados a um número pequeno de instituições, por exemplo, bibliotecas e museus.

Estas instituições normalmente utilizavam metadados para catalogar listas de objetos relacionados ao seu trabalho. Pensando na Web Semântica, pesquisadores desejam catalogar um número muito maior de recursos, distribuídos ao redor do mundo, registrados em diversos formatos e línguas de modo que possam ser utilizados e encontrados por qualquer pessoa ou agente de software.

Na essência o conceito de metadados é bastante simples, ou seja, são dados sobre dados. Quando se trata do mundo digital, chama-se de recurso o objeto descrito por metadados, pois este pode ser tanto



um simples dado, quanto um documento, uma página da web, ou até mesmo uma pessoa, uma coleção de livros, um sistema, um equipamento ou uma organização.

Metadados descrevem os recursos da web com a finalidade de facilitar a sua descoberta, localização e utilização. Motores de busca, ao utilizarem estes metadados, proporcionam consultas bem mais precisas, envolvendo não somente palavras, mas propriedades descritas, como o autor do recurso, o formato do recurso, a data do recurso, entre outros.

No mundo digital, metadados não se restringem a aqueles usados para descrever e localizar recursos. Podem também descrever os formatos dos recursos, a fim de permitir que computadores identifiquem quais aplicativos podem ser usados para manipular estes recursos (BREITMAN, 2006)

Segundo Milstead e Feldman (1999), em computadores os metadados servem para representar os recursos e para caracterizar as informações na Web de modo que todos os usuários entendam o significado das informações, ou seja, para que serve, a sua origem e suas condições de uso. A diversidade de tipos de metadados promove suporte à realização de inúmeras funções, tais como, descoberta, localização, avaliação, e uso da informação.

## 2.2 INFRAESTRUTURA

Para trabalhar com informações e para que seja possível disponibilizar os mesmos em formato aberto, é necessária a utilização de vários recursos para que se possa trabalhar com estes dados de forma correta e adequada. Nas próximas seções serão apresentadas algumas das ferramentas e padrões que devem ser utilizados para a disponibilização de dados abertos.

### 2.2.1 RDF

O RDF (*Resource Description Framework*) é uma linguagem voltada a representação de informações sobre recursos disponíveis na Web (W3C, 2004). O objetivo principal do RDF é representar uma codificação alinhada a um mecanismo de interpretação, para que *softwares* específicos para leitura destes dados sejam capazes de entender o conteúdo (ANTONIOU; HARMELEN, 2008). Além de ser uma das principais linguagens hoje para a Web Semântica, a grande maioria das representações de alto nível dependem ou resultam do RDF. Normalmente o RDF se utiliza da XML em sua semântica e URI (*Uniform Resource Identifier*) para a identificação de conceitos e entidades. URIs são rótulos utilizados para identificar um recurso abstrato ou físico na Web (BREITMAN; CASANOVA; TRUSZKOWSKI, 2007).

O RDF pode colaborar de forma positiva para a interoperabilidade entre os mais variados tipos de sistemas de informação e de descrição existentes atualmente, contribuindo assim, na construção de mecanismos de busca mais robustos e integrados que permitirão a oferta de serviços mais especializados aos seus usuários.

Para permitir que diferentes aplicações possam processar conteúdo da Web é importante que haja um padrão para este conteúdo disponibilizado. De maneira similar à utilização do HTML (*HyperText Markup Language*) atualmente na Web, em Open Data esse padrão é o RDF. De modo geral, RDF provê um modelo de dados que é extremamente simples por um lado, mas estritamente adaptado para a arquitetura Web por outro (BIZER; HEATH; BERNERS-LEE 2009).

Outra característica importante do RDF é a sua diversidade de modelos de seriação. Embora XML seja a especificação original de seriação, muitos outros formatos estão disponíveis, entre eles, Turtles, N-Triples e RDFa. Isso significa que não é necessário escrever, nem transmitir RDF na sintaxe XML. Como afirmam Antoniou e Harmelen (2008), a sintaxe baseada em XML não é um componente necessário do modelo RDF. A razão disso reside no fato do RDF ser uma forma de descrever dados e estruturas, não de apresentá-los (AUER et al., 2007).

Alguns servidores RDF podem oferecer os dados sobre o mesmo objeto ou entidade em diversos formatos distintos, de acordo com a solicitação do cliente. Essa característica confere flexibilidade no modo como as aplicações interoperam os dados.

Para que seja possível entender um pouco mais sobre a sintaxe e de como funciona o modelo de dados RDF torna-se necessário conhecer a estrutura interna deste modelo. Este modelo será utilizado quando for necessário descrever sobre algum recurso como um objeto, algum tipo de informação que se deseja falar ou descrever. Recursos podem ser autores, livros, editoras, lugares, pessoas, hotéis, quartos, consultas de pesquisa e assim por diante. Cada recurso tem um URI, podendo ser um URL (*Uniform Resource Locator*) ou algum outro tipo de identificador exclusivo. Nota-se que um identificador não necessariamente permite o acesso a um recurso. Esquemas URI foram definidos não só para locais de Web, mas também para objetos diversos como números de telefone, números de ISBN e localizações geográficas (ANTONIOU; HARMELEN, 2008).

RDF é um modelo de dados que é gerado em forma de triplas “Sujeito, Predicado e Objeto” e, desta forma simples, os dados podem ser facilmente representados e construídos em estruturas muito mais complexas. O sujeito de uma tripa é o URI do recurso identificado. O objeto pode ser o URI de outro recurso relacionado ao sujeito, ou mesmo pode ser um valor literal, como um texto, um número ou uma data. O predicado é sempre um URI que identifica o tipo de relacionamento entre o sujeito e o objeto. Os URIs desses predicados

vêm de vocabulários, coleções de URIs que podem ser usadas para representar informações sobre um certo domínio (BIZER; HEATH; BERNERS LEE, 2009).

Este modelo constituído em triplas serve tanto para representar os dados que são estruturados ou semi-estruturados. RDF é altamente significativo e essa capacidade de fácil expressividade permite definir vocabulários com semântica exata. Isso faz do RDF uma poderosa ferramenta para interoperabilidade e associação de dados e, fazendo uso desta capacidade, permite representar dados de maneira organizada e homogênea. Devido a sua flexibilidade, RDF pode ser visto como framework, modelo de dados, ou como vocabulário. Isso demonstra que o RDF é simples e complexo ao mesmo tempo (AUER et al., 2007).

A base para se ter uma Web de dados organizada são as ligações RDF, pois elas representam as ligações entre dois recursos. Uma ligação RDF consiste de uma tripla onde, sujeito e objeto, são os recursos interligados por um outro URI que identifica o tipo de relação entre ambos. Essas descrições contêm outras ligações RDF que, por sua vez, consistem-se de outros URIs que também podem ter seus dados retornados, e assim por diante. É assim que se pode navegar na Web de Dados, de forma semelhante ao que temos na Web clássica, que atualmente se utiliza de documentos ao invés de dados (BIZER; HEATH; BERNERS-LEE, 2009).

### 2.2.1.1 A origem do RDF

Segundo Miller (1998), o RDF teve início pela iniciativa do desenvolvimento do padrão (*Platform for Internet Content Selection*) PICS, padrão que serviria para auxiliar a Web na identificação de páginas que possuíssem conteúdo ofensivo, por exemplo. Isto possibilitaria que os pais ou responsáveis em casa ou em escolas pudessem configurar as páginas passíveis de serem acessadas.

Contudo, no decorrer do andamento do projeto foi percebido que seria necessário fornecer informações mais abrangentes sobre as páginas acessadas, não apenas algumas palavras. A questão da descrição de recursos se tornou muito mais ampla do que a proposta inicial do padrão PICS. A partir disto, criou-se um grupo de trabalho onde foi chamado de *Resource Description Framework* (RDF), onde o foco era desenvolver uma estrutura com as descrições dos recursos de modo abrangente para colaborar com as necessidades das várias comunidades de descrição interessadas (FERREIRA, 2013).

Várias linguagens e vocabulários de descrição influenciaram o RDF. Pode se dizer que em 1990, o que influenciou a criação do RDF foi a comunidade de descrição Dublin Core e principalmente a linguagem XML (MILLER, 1998).

Mas foi em 1999 que a W3C anunciou oficialmente o RDF, ainda que nesta época o objetivo fosse apenas a publicação e organização dos metadados e não a descrição de recursos em geral, o que aconteceu a partir de 2004 (W3C, 2004). Nesta época a W3C tinha planos em conseguir disponibilizar uma maior interoperabilidade na Web em que o objetivo principal era fornecer um padrão de dados abertos para que fosse possível descrever os recursos. Sendo assim o padrão RDF possibilitou um grande aumento no desenvolvimento de aplicações, que permitiam a realização de declarações a respeito de quase todo o tipo de objeto, e isso seria possível desde que este objeto pudesse ser identificado a partir de um URI (RAMALHO, 2006). Sendo possível descrever os recursos que podem não estar necessariamente disponível na Web, o RDF permite ser utilizado para realizar a troca de informações ou metadados nas mais diferentes áreas de aplicação, como é o caso da Biblioteconomia e mais especificamente podemos citar a catalogação de livros (FERREIRA, 2013).

#### 2.2.1.2 Modelo de dados RDF

RDF é um modelo de dados expresso em formato de triplas composto por Sujeito, Predicado e Objeto (BERGMAN, 2009). Alterando o Predicado por um verbo, e nomeando Sujeito e Objeto tem-se algo como: Pedro conhece Julia; ou a Roda é Redonda. Dessa forma simples os dados podem facilmente serem representados e construídos na Web em estruturas muito mais complexas. As triplas são chamadas de declarações (*statements*); Sujeito e Objeto são chamados de Nós e o predicado é chamado de Aresta.

As triplas são usadas para representar vários tipos de dados como os estruturados, semiestruturados e não estruturados. O modelo RDF é altamente representativo, e essa capacidade de representar permite definir vocabulários com semântica de acordo com a necessidade de determinado domínio. Isso torna o RDF uma poderosa ferramenta para ligação e associação de dados que podem ser originários de diversas fontes e com os mais diversos tipos de dados. Devido a sua flexibilidade, RDF pode ser visto como Modelo de Dados, ou como um tipo de Vocabulário. Isso mostra que RDF pode ser simples ou muito complexo dependendo do contexto (BERGMAN, 2009).

Segundo Miller (1998), o RDF, desenvolvido pelo W3C, possibilita uma infraestrutura em que desenvolvedores tenham condições de criar uma codificação voltada ao reuso de metadados estruturados. Esta infraestrutura permite a interoperabilidade de metadados por meio da concepção de mecanismos que suportam convenções comuns de semântica, de sintaxe e de estrutura. Ainda de acordo com Miller (1998), o RDF não denomina uma semântica para cada comunidade de descrição de recursos, mas sim, oferece a

capacidade para essas comunidades definirem os elementos de metadados conforme as suas necessidades específicas de descrição.

Para se criar estas estruturas pode ser utilizado o XML como sintaxe de escrita e processamento de dados. Através do XML é possível criar um modelo não ambíguo da semântica visando a consistência e a padronização dos dados.

Além desse novo formato de dados e semântica, Miller (1998), fala também de um modelo de RDF com o objetivo de facilitar a interoperabilidade entre os conjuntos de metadados. Isto inclui novos mecanismos padronizados para a representação de semântica, através de um simples, mas poderoso modelo de dados, o que é chamado de RDF Schema. RDF Schema pode ser visto como uma extensão no modelo RDF de modo a aumentar a legibilidade de vocabulários tanto para humanos quanto para máquinas.

Fazendo uso destes vocabulários, utilizando as URIs, o RDF é um modelo de dados formal que é utilizado para realizar a descrição dos recursos, deste modo é possível verificar que seu objetivo consiste em possibilitar que as mais variadas aplicações computacionais consigam trocar meta informação enquanto possam guardar seu significado original (FURGERI, 2006).

Recentemente o modelo de dados RDF tem se tornado importante perante aos olhos das comunidades de descrição de recursos. Como exemplo, pode ser citada a iniciativa da Dublin Core, comunidade responsável por promover a adoção de padrões de interoperabilidade de metadados e desenvolver vocabulários especializados para descrever fontes que possibilitem o desenvolvimento de sistemas mais inteligentes (BECKETT; MILLER; BRICKLEY, 2002).

### 2.2.1.3 Estrutura do RDF

O modelo de RDF é descrito em forma de grafos por meio de diagramas. Sendo assim o modo de representar os dados em RDF pode facilitar o aprendizado do modelo e a leitura por humanos, além disso, serve como um modo preciso para a criação de um modelo conceitual de determinado domínio.

Atualmente computadores não possuem adequada capacidade para o processamento de grafos considerando grandes volumes de informação (FERREIRA, 2013). Mesmo para humanos o entendimento de grafos somente é possível considerando pequenas estruturas, com poucos nós, arestas e literais. Mas os gigantescos bancos de dados atuais contêm milhares de dados que gerariam grafos absurdamente grandes. Para representar um conjunto de dados modelados em RDF, existem meios que não se utilizam de diagramas, mas sim de cadeias

de caracteres que podem ser processadas por máquinas, como podemos ver nos exemplos abaixo.

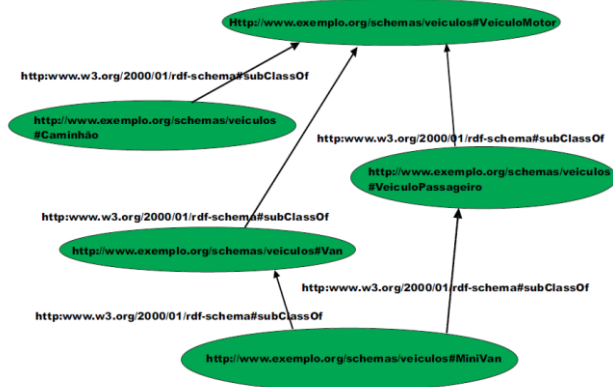
No exemplo abaixo (Quadro 1) é apresentado um RDF onde consta se define o tipo de cada veículo e sua classificação. A mesma informação pode ser vista na forma de um grafo (Figura 1).

Quadro 1 - Exemplo RDF sobre classificação de carros.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xml:base="http://www.exemplo.org/schemas/veiculos/">
<rdfs:Class rdf:ID="VeiculoMotor"/>
<rdfs:Class rdf:ID="VeiculoPassageiro">
<rdfs:subClassOf rdf:resource="#VeiculoMotor"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Caminhão">
<rdfs:subClassOf rdf:resource="#VeiculoMotor"/>
</rdfs:Class>
<rdfs:Class rdf:ID="Van">
<rdfs:subClassOf rdf:resource="#VeiculoMotor"/>
</rdfs:Class>
<rdfs:Class rdf:ID="MiniVan">
<rdfs:subClassOf rdf:resource="#Van"/>
<rdfs:subClassOf rdf:resource="#VeiculoPassageiro"/>
</rdfs:Class>
</rdf:RDF>
```

Fonte: Autor.

Figura 1 - Exemplo de grafo de classificação de carros.



Fonte: Autor.

### 2.2.1.4 Namespace

Como os nomes dos elementos em RDF não estão pré-definidos, isso pode levar a conflitos de nomes quando um documento RDF for agrupado com outro documento RDF, pois estes documentos podem usar os mesmos nomes para descrever tipos de elementos que não possuem relação um com o outro. Neste contexto o conceito de *namespace* é utilizado.

O *namespace* é um padrão definido pela W3C e serve basicamente para evitar os conflitos de nomes que são utilizados nos documentos em formatos XML, RDF e outras. A correção destes conflitos pode ser realizado se utilizados prefixos ou *namespace* (CAMATA; GOBIRA; MELOTTI, 2011).

Na Figura 2 apresenta informações que são referenciadas através do *namespace* “biglynx”. Figura Para tal, URIs são usadas para descrever pessoas e as relações entre elas fornecendo assim um contexto para os itens que são representados (nomes, termos técnicos, conceitos, entre outros), e que fornece diferenças para itens que possuem o mesmo nome, mas que residem em espaços de nomes diferentes. Tendo em vista que é fornecido um contexto distinto para cada *container*, o significado de um nome pode variar de acordo com o *namespace* o qual ele pertence.

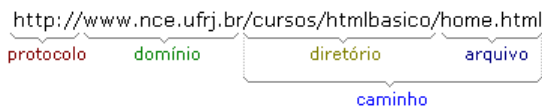
Figura 2 - URIs usadas para descrever pessoas e as relações entre elas.



Fonte: Heath (2011).

Deste modo é possível, a partir do exemplo da Figura 2, saber o que são e quem são os objetos descritos pelos URIs. O *namespace* nesse caso não se restringe apenas ao domínio *biglynx.co.uk*, mas engloba também o caminho subsequente *people*, ou seja, o *namespace* do exemplo é “*http://biglynx.co.uk/people*”. Com esse espaço de nomes pode-se contextualizar, considerando a figura, quais objetos são pessoas e pertencem ao domínio ou tem algum vínculo com a “Big Lynx”. Percebe-se, portanto, que é possível nesse formato identificar facilmente os objetos, de forma a permitir seu rápido reconhecimento, bem como o tratamento da ambiguidade de termos que representam os objetos presentes em determinado contexto. A Figura 3 apresenta o esquema da estrutura de um URI.

Figura 3 - Exemplo de endereço de URI.



Fonte: NCE (2013).

### 2.2.2 Jena

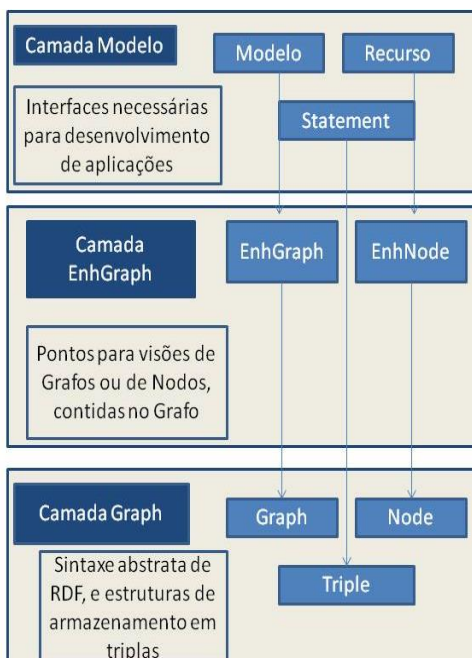
Jena foi originalmente desenvolvido por pesquisadores da HP Labs em Bristol, Reino Unido, no ano de 2000. É um projeto *open-source* e tem sido amplamente utilizado em uma variedade de aplicações voltadas à Web Semântica. Em 2009 a HP decidiu ceder o projeto à organização Apache visando alavancar novos desenvolvimentos e em 2010, a Apache adotou o projeto Jena para trabalhos futuros. Neste mesmo ano, em novembro, o projeto entrou em incubação e vem sendo evoluído desde então (APACHE, 2011).

Jena é um *framework* na linguagem Java para programação de aplicações para Web Semântica, o qual provê ambientes para programação de RDF, RDF Schema (RDFS) e OWL, além de incluir mecanismos de inferência baseado em regras. O *Framework* Jena como mencionado acima foi desenvolvido pela HP Labs com o objetivo de tornar fácil o desenvolvimento de aplicações que usem linguagens e modelos baseados na Web Semântica. O Jena 1 foi lançado em 2000 e dentre suas diversas contribuições destacam-se as ferramentas para manipulação de grafos RDF e o suporte de ontologias através de DARPA *Agente Markup Language + Ontology Inferency Layer*(DAML+OIL), enquanto o Jena 2 foi lançado em 2003 e se caracterizou pelo suporte a RDFS e OWL, além de prover mecanismos de inferência para ambas as linguagens (CARROLL et al., 2003).



A arquitetura (Figura 4) é organizada em uma estrutura dividida em camadas, onde em sua camada principal encontra-se a *Graph*, baseada da sintaxe abstrata de RDF e provendo estruturas de armazenamento de triplas (Sujeito+Predicado+Objeto). Esta camada também fornece os mecanismos de inferência para RDF e OWL. A segunda camada é a *EnhGraph*, a qual se caracteriza por ser uma camada intermediária com extensões para visões de grafos e nodos que estão contidos em grafos. A última camada é chamada de *Model* e disponibiliza aos programadores as *interfaces* necessárias para o desenvolvimento de aplicações (CARROLL et al., 2003).

Figura 4 - Arquitetura Jena.



Fonte: Adaptado de Carrol et al. (2003).

### 2.2.2.1 Criação de modelos

Para a criação de modelos para triplificação de dados é necessário definir a forma de armazenamento dos dados, sendo, na memória, em bancos de dados relacionais (SDB) ou na versão disponibilizada nativamente pelo Jena (TDB). Para tal, faz-se necessário a utilização da API de modo que aplicações possam ser desenvolvidas com o intuito de armazenar e consultar conteúdo em RDF.

O Quadro 2 apresenta um exemplo de método de criação do modelo em um banco de dados Postgres® para triplicação em uma base de dados relacional.

Quadro 2 - Método para criação de modelo.

```
private IDBConnection conn;
private ModelMaker maker;
private Model model;
private String M_DB_MODEL;
....

private void startModel() throws Exception {
    if (!conn.containsModel(M_DB_MODEL)) {
        maker = ModelFactory.createModelRDBMaker(conn);
        model = maker.createModel(M_DB_MODEL, true);
    } else {
        maker = ModelFactory.createModelRDBMaker(conn);
        model = maker.openModel(M_DB_MODEL, true);
    }
}
```

Fonte: Autor.

No exemplo acima é verificado se o modelo já está criado ou não no banco de dados através do método `containsModel()` da classe `IDBConnection`. Caso o modelo no banco de dados não exista, ao se invocar o método `createModel()` de `ModelFactory`, as tabelas necessárias para o armazenamento do conteúdo RDF são criadas. Entre as tabelas existe uma de maior relevância, pois contém três colunas principais que possibilitam representar o Sujeito, o Predicado e o Objeto de uma tripla RDF. Nas duas situações, criação ou abertura com sucesso, será instanciado um objeto da classe `Model` que viabiliza a submissão de comandos em linguagem adequada para a manipulação das informações no banco de dados.

### 2.2.3 SPARQL

Assim como os sistemas de bancos de dados relacionais fazem uso da linguagem de consulta SQL para consultar registros nas suas bases de dados, o SPARQL é uma linguagem para consulta em formato RDF que se tornou padrão com recomendação da W3C em 2008. É uma linguagem orientada a dados que recupera as informações contidas em arquivos RDF, possibilitando inclusive a opção de combinar dados de arquivos que se encontram em diferentes locais da Web (W3C, 2008). À medida que mais dados passaram a ser

disponibilizados em formatos RDF surgiu a necessidade de uma maneira simples de localizar informações específicas. Surge então a SPARQL, uma linguagem de consulta com a finalidade de facilitar a localização de dados em formato RDF ou em bases triplicadas (IBM, 2005).

SPARQL pode ser usado para realizar consultas através de diversas fontes de dados, não importando se os dados estão nativamente em RDF ou se são somente visualizados como RDF através de uma camada auxiliar. Suporta também testes de valores extensíveis e restrição de consultas por fonte do grafo RDF. Os resultados das consultas podem ser um conjunto de resultados com diversas linhas com os valores em colunas, similar ao resultado de consultas SQL ou na forma de grafo RDF (W3C, 2008).

A linguagem de consulta SPARQL é amplamente utilizada na consulta de dados RDF e é implementada por várias empresas que oferecem soluções para manipulação de conteúdo no padrão RDF (TAUBERER, 2006).

A linguagem SPARQL segue a mesma estrutura de construção de arquivos RDF e é construída sobre padrões de triplas (*triple pattern*), ou seja, Sujeito (Subject), Predicado (Predicate) e Objeto (Object).

### 2.2.3.1 Exemplos de consultas em SPARQL

Uma consulta SPARQL é composta por um conjunto de cláusula, sendo que as cláusulas básicas se assemelham ao comando “SELECT” da linguagem SQL. O Quadro 3 apresenta a estrutura básica.

A cláusula SELECT especifica a projeção, ou seja, identifica quais valores serão apresentados. Diferentemente da SQL que projeta determinada coluna de uma tabela ou visão, em SPARQL o valor que preenche determinada variável criada durante a consulta será apresentado. A cláusula FROM possibilita identificar qual fonte de dados será consultada, neste caso, um arquivo RDF disponível em determinado URL. Esta cláusula é opcional uma vez que a fonte de dados pode estar em memória ou em banco de dados. Por fim, a cláusula WHERE permite definir os padrões de triplas/grafos a serem localizados na base RDF.

Quadro 3 - Cláusulas básicas da linguagem SPARQL.

SELECT ... FROM ... WHERE { ... }
---

Fonte: Autor.

A seguir são apresentadas algumas possibilidades de consulta SPARQL. O Quadro 4 apresenta uma consulta básica em que, através da cláusula WHERE, são selecionados todos os títulos de livros em que o Sujeito seja igual a “<http://example.org/book/book1>” e o predicado igual a <http://purl.org/dc/elements/1.1/title>. Todas as triplas que satisfazem tal condição terão o valor da propriedade “title” atribuído para a variável “?title” que será então projetada através da cláusula SELECT.

Quadro 4 - Consulta SPARQL básica.

```
SELECT ?title WHERE {
  <http://example.org/book/book1>
  <http://purl.org/dc/elements/1.1/title> ?title . }
```

Fonte: Autor.

Considerando a consulta anterior nota-se que tanto a informação do Sujeito quanto do Predicado envolve a declaração completa de um URI. Isto dificulta o entendimento e eventuais manutenções e evoluções nas consultas. Para lidar com isto a linguagem SPARQL permite a inserção de prefixos para declarar um esquema usado na consulta. O Quadro 5 apresenta um consulta utilizando dois prefixos, um chamado “cidade” e outro chamado “estado”. Pode-se notar que a consulta fica mais clara e, portanto, mais fácil de ser manipulada.

Quadro 5 - Exemplo de uso de Prefixo.

```
PREFIX cidade:<http://DadosAbertos.org/cidades#>
PREFIX estado: <http://DadosAbertos.org/estados#>
SELECT ?Codigo_Cidade ?Nome_Cidade ?Nome_estado
WHERE
{
  ?x cidade:i_cidades ?Codigo_Cidade .
  ?x cidade:nome ?Nome_Cidade .
  ?x cidade:i_estados ?Sigla_partido .
  ?x cidade:i_estados ?i_estado .
  ?i_estado estado:nome ?Nome_estado .
}
```

Fonte: Autor.

O exemplo apresentado no Quadro 6 indica que parte do padrão de consulta é opcional. Neste caso, serão apresentados os conteúdos das variáveis que representam o primeiro nome e a idade para todas as triplas tipificadas como “Autor”. No caso da variável idade, o conteúdo desta somente será projetado se a propriedade “idade” existir para determinado “Sujeito” atribuído em “?x”, caso contrário, o valor será vazio.

#### Quadro 6 - Clausula OPTIONAL.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX bibl: <http://www.example.org/biblioteca#>
SELECT ?primNome ?idade WHERE{
  ?x rdf:type <bibl:Autor> .
  ?x bibl:primeiroNome ?primeiroNome .
  OPTIONAL { ?x bibl:idade?idade}
}
```

Fonte: Autor.

A cláusula UNION (Quadro 7) possibilita definir padrões alternativos em uma determinada consulta, no exemplo, a recuperação de adultos, previamente designados à classe “Pessoa\_Adulta” ou que possuam idade maior do que 17 anos. Na segunda parte da consulta ocorre a cláusula FILTER utilizada para restringir triplas a partir de teste de conteúdo de determinada variável.

#### Quadro 7 - Clausula UNION.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://www.example.org#>
SELECT ?primeiroNome
WHERE{
  ?x ex:primeiroNome ? primeiroNome .{
    { ?x rdf:type <ex:Pessoa_Adulta> }
    UNION
    { ?x ex: idade ?idade
      FILTER (?idade > 17) }}
}
```

Fonte: Autor.

Outras cláusulas também utilizadas em SQL convencional são a ORDER BY que possibilita ordenar o resultado, LIMIT para limitar a quantidade de itens que serão demonstrados e OFFSET que determina o deslocamento do conjunto resultante a partir do primeiro item (Quadro 8).

### Quadro 8 - Cláusulas ORDER BY, LIMIT e OFFSET.

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
FROM <http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>
WHERE {
  ?person foaf:name ?name .
}
ORDER BY ?name
LIMIT 20
OFFSET 20
```

Fonte: Autor.

A cláusula UNBOUND é utilizada para testar se uma variável da tripla está ou não instanciada, permitindo assim a negação por falha. No exemplo do Quadro 9 as triplas que não possuem idade serão recuperadas.

### Quadro 9 - Cláusula UNBOUND.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://www.example.org#>
SELECT ?firstName
WHERE
{
  ?x ex:firstName ?firstName .
  OPTIONAL { ?x ex:age ?age }
  FILTER ( !bound(?age) )
}
```

Fonte: Autor.

No Quadro 10 é apresentada a operação de negação que, no exemplo, irá recuperar pessoas que não possuem como habilidade a linguagem de programação Java®.

### Quadro 10 - A clausula de negação.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://www.pessoas.org#>
SELECT ?nome
WHERE {
  ?x ex:nome ?nome .
  ?x ex:habilidades ?habilidades.
  ?habilidades?y ?z .
  FILTER ( ?z != "Java" )
}
```

Fonte: Autor.

Por fim, o Quadro 11 apresenta a cláusula ASK. Por meio desta cláusula é possível realizar uma pergunta levando em consideração as demais cláusulas e que terá como retorno um valor positivo ou negativo, ou seja, se existem triplas que satisfazem determinada restrição ou não.

Quadro 11 - Cláusula ASK.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://www.pessoas.org#>
ASK {
  ?x ex:idade ?idade .
  FILTER ( ?idade > 25 )
}
```

Fonte: Autor.

## 2.3 POSSIBILIDADES DE DADOS ABERTOS

Atualmente muitas entidades governamentais e privadas vêm trabalhando com a disponibilização de dados abertos. De modo geral, espera-se que a disponibilização de dados abertos possa ajudar pessoas e organizações de alguma maneira. No Brasil o governo vem promovendo o primeiro passo nesta direção. Vários outros países possuem iniciativas de disponibilização de seus dados governamentais na Web.

Os dados abertos podem ajudar as pessoas a tomar decisões melhores no seu cotidiano ou permitir que sejam mais ativas na sociedade. Criado na Dinamarca, a aplicação *Find Toilet*, mostra em um mapa todos os sanitários públicos acessíveis que existem no país. Esta informação pode parecer sem importância alguma, mas possibilita que pessoas com problemas de saúde saiam de casa com mais frequência por não temerem a dificuldade de encontrar sanitários.

Na Holanda, o site *Vervuiling Alarm* avisa com uma mensagem se a qualidade do ar da vizinhança esta perto a atingir um nível muito prejudicial a saúde. Em Nova Iorque, uma ferramenta busca lugares onde é possível caminhar com um cachorro, como por exemplo, parques (GRAY et al., 2012).

Existem projetos como o Mapumental<sup>2</sup>, no Reino Unido, e o Mapnificent<sup>3</sup>, na Alemanha, que permitem encontrar locais para morar de acordo com características indicadas, como duração do trajeto até o

<sup>2</sup> <http://mapumental.channel4.com/>

<sup>3</sup> <http://www.mapnificent.net/>

local de trabalho, os valores das casas e a beleza da região. Estes são alguns exemplos de serviços que utilizam dados abertos.

Os Dados abertos também podem ter grande importância na economia. Por exemplo, o Husetsweb<sup>4</sup>, da Dinamarca, ajuda a descobrir novas formas para melhorar a eficiência energética da sua residência, incluindo uma ferramenta para realizar o planejamento financeiro e uma busca por empreiteiros que possam realizar o trabalho necessário para economizar mais.

Outro exemplo que vem crescendo muito é o Google Translate<sup>5</sup>. Este serviço aproveita a grande quantidade de publicações, principalmente quando disponibilizado em línguas européias para treinar seus algoritmos de tradução de modo que se possa melhorar a qualidade do serviço realizado (GRAY et al., 2012).

Dados abertos também são de grande importância para o próprio governo. Por exemplo, o Ministério da Educação da Alemanha disponibilizou na web, para que todas as pessoas pudessem utilizar e visualizar, todos os dados governamentais da educação no país.

Com esta iniciativa de disponibilizar estes dados, o número de perguntas que recebiam da população referente a dados da educação teve uma queda muito grande, reduzindo a carga de trabalho dos funcionários e custos do governo. Além disso, as perguntas que continuam sendo feitas tornaram-se mais simples de serem respondidas pelos servidores públicos. Isto se deve ao fato dos cidadãos saberem de forma mais exata sobre o que vão pedir, pois os dados estão disponíveis para todos e as pessoas podem ler e se inteirar sobre o assunto antes de realizar as perguntas para os servidores (GRAY et al., 2012).

Outra motivação para a utilização da publicação dos dados abertos são os impactos econômicos que eles podem gerar para as empresas, governo ou qualquer entidade que trabalhe com esse formato de disponibilização de dados. Esse aspecto pode despertar interesse principalmente nos governos e também de empresas privadas, no intuito de melhorar sua economia ou lucros no caso de uma empresa privada.

Com dados abertos novas oportunidades de negócio são criadas constantemente. A publicação de dados sobre trânsito em formatos amigáveis já são realidade em vários locais do país. De maneira geral, tal publicação possibilita vários serviços, tais como, a identificação de serviços próximos ao cidadão, programas para acompanhamento sobre ações do poder legislativo, informações precisas sobre imóveis à venda, como antigos donos, ano de construção, entre outros.

---

<sup>4</sup> <http://www.husetsweb.dk/>

<sup>5</sup> <http://translate.google.com.br/>



Os ganhos para o governo também são mensuráveis. Isto pode ser evidenciado, por exemplo, no aumento da arrecadação de impostos pela geração de emprego, criação de novas oportunidades de negócio, a criação de novas empresas e a ampliação de negócios fomentados pela agregação de bases de dados abertos públicos. Segundo Eaves (2010), a publicação de dados abertos permitiu ao governo canadense identificar 3.2 bilhões que haviam sido desviados dos cofres públicos por falsas organizações de caridade.

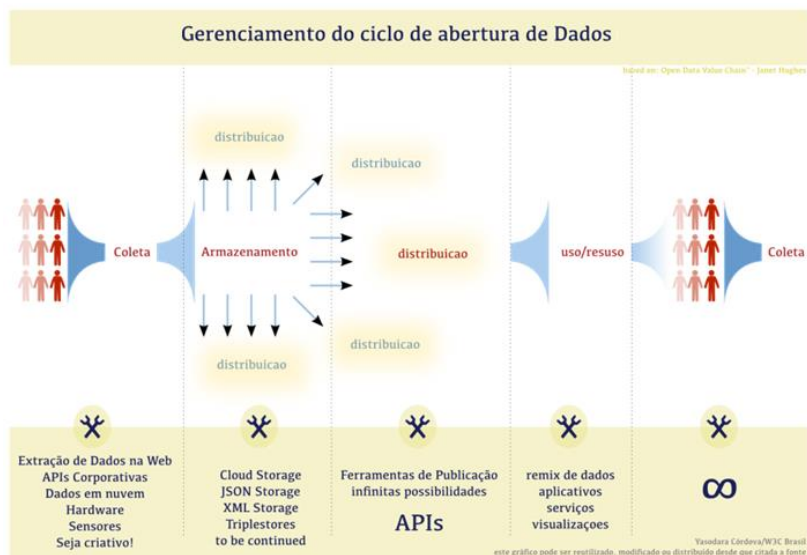
Segundo Vickery (2011), estima-se que os ganhos com o reuso de dados públicos na Europa tenha sido de 32 bilhões de Euros em 2010. Considerando que essas motivações são suficientes para a publicação de Dados Abertos Governamentais, é importante frisar que, para possibilitar que dados abertos sejam utilizados e representem alguma vantagem para as empresas e para os próprios publicadores, os dados precisam ter qualidade e sua disponibilidade precisa ser garantida a todos.

### 3 CENÁRIO DE APLICAÇÃO E MODELO PROPOSTO

Publicar dados abertos vai além de apenas disponibilizá-los de forma aberta e gratuita em qualquer formato. O esperado é que tais dados estejam disponíveis em formatos passíveis de processamento automatizado por computadores, e que existam informações sobre a definição e estrutura desses dados. Este capítulo apresenta o desenvolvimento do processo de publicação de um pequeno conjunto de dados do Site DivulgaCand em formato de dados abertos.

De acordo com a Figura 5, que ilustra as principais fases para a publicação de dados abertos, o ciclo que um determinado dado percorre começa pela "Coleta". Segundo a OD4D (2013), o atual estágio da sociedade, muito mais informatizada e tecnológica e com dispositivos de *hardware* disponíveis por preços acessíveis, possibilita uma maior criatividade no que tange a elaboração de aplicações que possam beneficiar pessoas.

Figura 5 - Ciclo de disponibilização de dados abertos.



Fonte: OD4D (2013).

Depois da coleta dos dados, independente da intenção, ou seja, publicação ou republicação torna-se necessário que as informações sejam armazenadas em repositórios estruturados e planejados para receber e distribuir os mesmos de forma aberta e que possibilitem a interoperabilidade.

A abertura destes dados é importante, pois permite ao cidadão conferir, analisar, acompanhar e sugerir políticas públicas que sejam mais adequadas às suas necessidades. O objetivo principal do ciclo de abertura de dados é permitir a conexão do Estado com os cidadãos de modo que este consiga coletar/extrair os dados ou permitir o uso e reuso desses dados (OD4D, 2013).

A partir deste contexto inicial é apresentado nas seções seguintes um detalhamento do cenário de aplicação, bem como, o modelo proposto com o detalhamento de seus componentes.

### 3.1 PORTAL DO CANDIDATO

Com o objetivo de disponibilizar os dados das candidaturas e visando mais transparência para os cidadãos brasileiros, a Lei No 8.730, de 10 de Novembro de 1993, exige que os candidatos declarem os seus bens à Justiça Eleitoral. Sendo assim, os eleitores podem pesquisar de maneira online as informações que os candidatos a prefeito e a vereador declararam ao Tribunal Superior Eleitoral (TSE). Todas essas informações estão disponíveis no site DivulgaCand 2012 (TSE, 2012). A Figura 6 apresenta a tela inicial do Site DivulgaCand.

Figura 6 - Pagina principal do Site DivulgaCand.



Fonte: TSE (2012).

O Tribunal Superior Eleitoral disponibilizou na internet o Site DivulgaCand 2012 em que todos os partidos políticos, coligações, candidatos e a população podem se informar sobre a quantidade e os dados dos candidatos que vão concorrer a prefeito, vice-prefeito e vereador nas Eleições de 2012 (TSE, 2012). Os dados são atualizados pela Justiça Eleitoral com dados agregados nacionalmente por Estado e por Municípios (Figura 7). Pelo sistema, as pessoas interessadas podem saber ainda o número de vagas para vereador que serão disputadas em cada município (TSE, 2012).

Por meio do Site DivulgaCand também é possível consultar o número de candidatos a prefeito e vereadores por município e por cargo. Pode ser visualizada também a situação do pedido de registro da candidatura, além de permitir a realização de pesquisas sobre as informações repassadas à Justiça Eleitoral como a declaração de bens dos candidatos, as certidões criminais e a previsão de gastos de campanha.

Figura 7 - Informações de candidatos disponibilizadas por estado e município.

**DivulgaCand**  
Divulgação de Registro de Candidaturas 2012

Selecione a situação do candidato e do município desejados para efetuar a pesquisa. Clique aqui para saber mais sobre as situações existentes.  
Situação do Candidato: ☐ Inativo ☐ Inapto ☒ Todas

Nome do Município	Candidatos a Prefeito	Candidatos a Vereador	Vagas para Vereador
ARARAQUÃ	5 R	113 R	15

[ 1 item(ns) encontrado(s). Mostrando de 1 a 1 ] (filtrado de 295 registros no total)

Filtrar candidatos para Vereador em ARARAQUÃ (SC)

Nome do Candidato	Nome para urna	Número	Situação	Partido	Coligação
ADAIR JORDAO	PROFESSOR JORDAO	15013	Defendido	PT	ARARAQUÃ PODE MUITO MAIS
ADAIR MENDES	PIRELAMPO	65123	Defendido	PC do B	Araraquã Pode Mais
ALEXANDRE REZENDE PEREIRA	ALEXANDRE PEREIRA	23123	Defendido	PPS	Araraquã Pode Mais
AMILTON INO MAZARIO	AMILTÃO	55555	Defendido	PSD	PARA ARARAQUÃ CONTINUAR CRESCENDO
ANDERSON DA SILVA	ANDI	12323	Defendido	POT	ARARAQUÃ PODE MUITO MAIS
ANDREIA JAINE CONSTANTINO GONÇALVES	ANDREIA CONSTANTINO	13413	Indeferido	PT	ARARAQUÃ PODE MUITO MAIS
ANTONIO CARLOS ORIGI	TORINHO ENCANADOR	15500	Defendido	PMDB	Perdido não coligado
ANTONIO PEREIRA	TORINHO DA ROÇA	12810	Defendido	POT	ARARAQUÃ PODE MUITO MAIS
ANTÔNIO ROBERTO LEANDRO	BETO DO POSTO DE SAUCE	23666	Defendido	PPS	Araraquã Pode Mais
AQUILES GHELLERE	KILA	40123	Defendido	PSB	ARARAQUÃ PODE MUITO MAIS
ARILTON DE SOUZA COSTA	ARILTON COSTA	11444	Defendido	PP	PARA ARARAQUÃ CONTINUAR CRESCENDO
CELDO DE SOUZA	PROFESSOR CELDO	16116	Defendido	PSTU	FRENTE DE ESQUERDA ARARAQUÃ
CLAUDIA GOLLART	CLAUDIA GOLLART	11200	Defendido	POT	ARARAQUÃ PODE MUITO MAIS
CLAUDIO ROBERTO DOS PASSOS	CLAUDIO ROBERTO	12660	Defendido	POT	ARARAQUÃ PODE MUITO MAIS
CLAUDIONEI PEREIRA BORGES	NEI DA METALÚRGICA	20100	Defendido	PSC	Araraquã Pode Mais

[ 113 item(ns) encontrado(s). Mostrando de 1 a 15 ]

Exportar dados de candidatos  
(\*) Eleições suplementares

Primeiro Anterior 2 3 4 5 Próximo Último

**SC - SANTA CATARINA**  
Quantidade de pedidos de registros de candidatura efetuados até o momento:

Prefeito: 714  
Vice-Prefeito: 726  
Vereador: 16254

Última atualização dos dados  
10/07/2012 06:01:34  
Versão: 2012-1.1-10

Acesso a Estatísticas de Candidaturas

Selecione o TSE

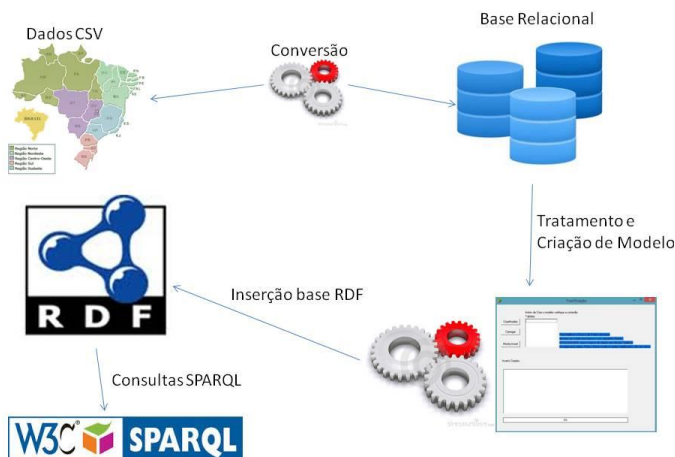
Fonte: TSE (2012).

### 3.2 MODELO PROPOSTO

A Figura 8 apresenta o modelo proposto e detalha as etapas que foram realizadas para viabilizar a conversão dos dados do Site DivulgaCand em um modelo de dados abertos em formato RDF.

Inicialmente, o modelo prevê a obtenção dos dados em formato disponibilizado pelo Site DivulgaCand. Após a coleta dos dados os mesmos são convertidos para um banco de dados. Para tal, foi proposto um modelo de dados relacional, que será discutido nas próximas seções, visando representar o espaço de dados que possivelmente suporta o Site DivulgaCand. Vale mencionar que este passo somente se justifica uma vez que os dados para o cenário escolhido não estão em formato de dados aberto padronizado.

Figura 8 - Modelo proposto para a disponibilização dos dados do Site DivulgaCand.



Fonte: Autor.

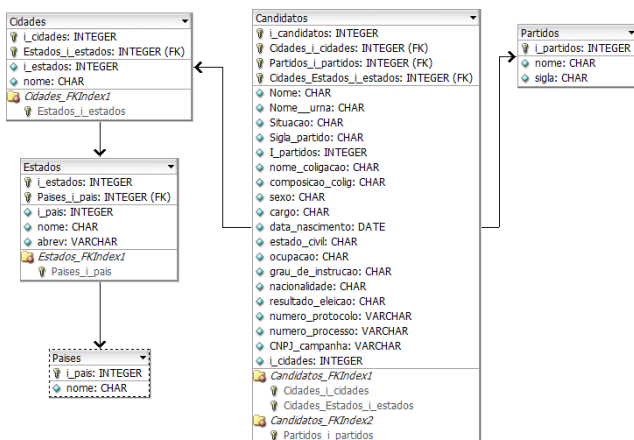
A próxima etapa envolve a transferência dos dados relacionais para o formato RDF armazenado em um banco de dados capaz de suportar este formato. Para tal, foi desenvolvida uma ferramenta, detalhada nas próximas seções, para facilitar o processo de transferência.

A última etapa objetiva a seleção das informações no banco de dados de triplas através de consultas na linguagem SPARQL. Através dessas consultas, torna-se possível que desenvolvedores criem aplicações com diferentes finalidades.

### 3.2.1 Modelo Relacional

Os dados do Site DivulgaCand são disponibilizados em formato CSV, ou seja, formato de texto em que cada campo é separado por determinado caractere. Cada linha desse arquivo foi inserida em tabela correspondente por meio de comandos SQL (*Structured Query Language*) de inserção. Neste trabalho propõem-se um modelo de dados capaz de representar minimamente o Site DivulgaCand (Figura 9). Esta abstração tem o intuito de demonstrar como poderia ser o modelo de dados uma vez que não se teve acesso ao modelo original. Além disso, pensando em dados abertos, os dados poderiam ser disponibilizados em formato RDF, ao invés de CSV, facilitando integração com outras aplicações.

Figura 9 - Modelo Entidade-Relacionamento proposto para suportar os dados originais do Site DivulgaCand.



Fonte: Autor.

O modelo sugerido é composto de cinco tabelas em que a tabela central “Candidatos” armazena as informações básicas do candidato, tais como, nome, a cidade, o estado e o partido pela qual este irá concorrer. As demais tabelas, “Partidos”, “Cidades” e “Estados” são auxiliares ao processo e mais voltadas à normalização dos dados.

### 3.2.2 Ferramenta de Conversão

Com o intuito de facilitar a tarefa de transferência da base relacional para a base triplificada foi desenvolvido um protótipo em Delphi para criação de modelos e base de dados no formato de triplas.

Com essa ferramenta é possível extrair dados de qualquer banco de dados relacional e transformar em um banco de dados no formato de triplas. Para que isto seja possível, a aplicação analisa o metadados do esquema do banco de dados de origem visando identificar as tabelas e as colunas de cada tabela. A partir disto, seleciona todas as linhas fazendo a migração para a base triplificada. O detalhamento do protótipo será realizado na seção 4.1.

### 3.2.3 Modelo RDF

Como mencionado, para que a migração seja possível é necessário que a aplicação liste todas as tabelas e suas colunas no modelo de origem. Isto é realizado inspecionando o metadados. Além disto, é também necessário transformar o conteúdo de uma tabela

relacional em triplas RDF, ou seja, transformar todo o banco de dados selecionado em uma estrutura composta de Sujeito, Predicado e Objeto.

Após os dados serem selecionados e devidamente convertidos em triplas é necessário importar os mesmos no banco de dados destino. O Quadro 12 apresenta o resultado em RDF para um determinado candidato.

Quadro 12 - Exemplo em RDF/XML de um candidato e demais informações relacionadas.

```
<?xml version="1.0"?>
<!DOCTYPE rdf:RDF [<!ENTITY xsd
"http://www.w3.org/2001/XMLSchema#">]>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:cand="http://Divulgacand.org/Divulgacand/Candidatos#"
  xmlns:cid="http://Divulgacand.org/Divulgacand/Cidades#"
  xmlns:est="http://Divulgacand.org/Divulgacand/Estados#"
  xmlns:part="http://Divulgacand.org/Divulgacand/Partidos#"

  <rdf:Description rdf:about="cand:1">
    <cand:nome>Candidato do Partido Tal</cand:nome>
    <cand:nome_urna>Candidato</cand:nome_urna>
    <cand:situacao>Apta</cand:situacao>
    <cand:i_partidos rdf:resource="part:100"/>
    ....
    <cand:cargo>Vereadora</cand:cargo>
    <cand:sexo>Feminino</cand:sexo>
    <cand:data_nascimento>12/12/1980</cand:data_nascimento>
    <cand:i_cidades rdf:resource="cid:4214409"/>
  </rdf:Description>
  <rdf:Description rdf:about="cid:4214409">
    <cid:nome>Rio das Antas</cid:nome>
    <cid:i_estados rdf:resource="est:24"/>
  </rdf:Description>
  <rdf:Description rdf:about="est:24">
    <est:nome>Santa Catarina</est:nome>
    <est:abrev>SC</est:abrev>
  </rdf:Description>
  <rdf:Description rdf:about="part:100">
    <part:nome>Partido Progressista</part:nome>
    <part:sigla>PP</part:sigla>
  </rdf:Description>
</rdf:RDF>
```

Fonte: Autor.

Pensando no formato de tripa RDF cada candidato é caracterizado como um recurso identificado por um URI e composto de

várias propriedades. O URI designa o Sujeito, uma propriedade em particular determina o predicado e o valor da propriedade o Objeto. Em geral, existirão tantas triplas quanto forem as propriedades de um recurso.

Além da representação em RDF/XML a mesma informação pode ser visualizada na forma de grafo. A Figura 10 apresenta as mesmas informações do exemplo do candidato em RDF em que o URI é identificado como o nó principal, a esquerda representando o Sujeito e conectado a vários outros nós a direita (Objetos) através de linhas (arestas) que representam os Predicados.

Figura 10 - Exemplo de grafo para a informação de um candidato em particular.



Fonte: Autor.

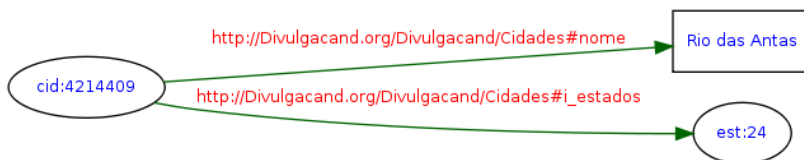
Também pode ser observado na Figura 10 de forma mais clara como funciona uma tripla. É possível identificar mais detalhadamente o Sujeito, Predicado e Objeto, em que o Sujeito designado pelo URI



“<http://Divulgacand.org/Divulgacand/Candidatos/1>” está ligado pelos predicados correspondentes. Pensando em uma tabela relacional estes predicados são representados pelas colunas adicionais a chave primária. No exemplo, a coluna da tabela referenciada pelo Sujeito é a *i\_candidatos* que possui como conteúdo o valor ‘1’.

Considerando o modelo relacional existem informações complementares que descrevem as chaves estrangeiras. Na tradução para RDF a informação (Objeto) é identificada através de um URI. A partir deste ponto o Objeto transforma-se em um Sujeito que é então descrito através de novas propriedades que identificam o recurso. Na Figura 11 consta o exemplo da cidade “Rio das Antas” que possui o URI “<http://Divulgacand.org/Divulgacand/Cidades/4214409>” e as propriedades *nome* e *i\_estados* que aponta para o detalhamento do estado.

Figura 11 - Exemplo de grafo para Cidade.



Fonte: Autor.

Após a etapa de carga das informações em um banco de dados triplificado torna-se possível a seleção destas mesmas informações utilizando-se a linguagem SPARQL.

## 4 DESENVOLVIMENTO DO PROTÓTIPO E AVALIAÇÃO DOS RESULTADOS

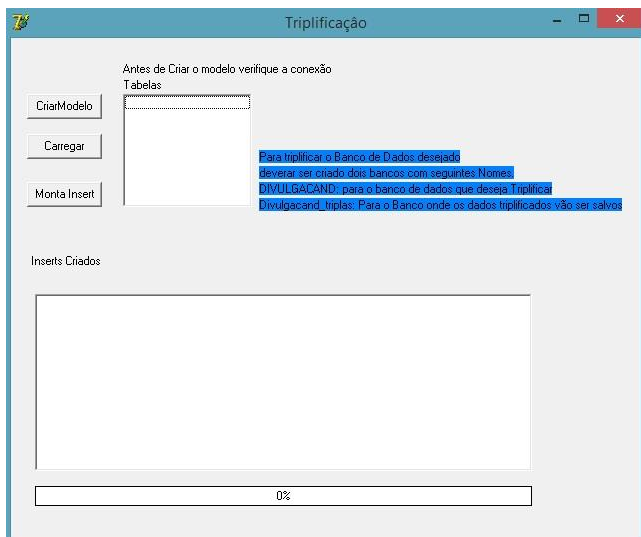
Nesta seção será detalhado o protótipo desenvolvido e serão apresentadas algumas consultas realizadas sobre a base de dados abertos com o objetivo de avaliar a consistência das informações disponibilizadas em formato padronizado.

### 4.1 DETALHAMENTO DO PROTÓTIPO

O protótipo foi desenvolvido com algumas funcionalidades básicas para realizar a carga de um banco de dados no formato de triplas. Para tal, utilizou-se o *framework* Jena® disponível na linguagem Java® que possibilita a criação do modelo RDF em um Banco de Dados. Neste trabalho utilizou-se o banco de dados Postgres®. A interface e as funcionalidades do protótipo foram implementadas utilizando as linguagens de programação Java® e Delphi®.

A Figura 12 apresenta a tela principal do protótipo desenvolvido. O processo de transferência das informações do modelo relacional para o modelo triplificado ocorre em três passos, sendo: a) a criação do modelo triplificado (botão “Criar Modelo”); b) a carga dos metadados do modelo de origem (botão “Carregar”); e c) a transferência dos dados entre os modelos (botão “Monta Insert”).

Figura 12 - Protótipo desenvolvido para possibilitar a migração dos dados.



Fonte: Autor.

No primeiro passo, o protótipo realiza a conexão com os bancos de dados de origem e de destino, onde os dados da fonte de origem serão migrados e transformados em modelo de triplas no banco de destino. Também neste momento, através do *framework* Jena®, é criado o modelo triplicado no banco de dados Postgres®.

Nos Quadros Quadro 13 e Fonte: Autor.

Quadro 14 a seguir é possível verificar os métodos responsáveis pela criação e fechamento das conexões com os bancos de dados, origem e destino, permitindo o início do processo de migração dos dados relacionais para dados em triplas RDF.

Quadro 13 - Método utilizado para estabelecer as conexões com os bancos de dados de origem e destino.

```
private void openDB() throws Exception {
    Class.forName(M_DBDRIVER_CLASS);
    sourceConn = DriverManager.getConnection(M_DB_URL_1,
        M_DB_USER,M_DB_PASSWD);
    targetConn = new DBConnection(M_DB_URL_2, M_DB_USER,
        M_DB_PASSWD, M_DB);
}
```

Fonte: Autor.

Quadro 14 - Método utilizado para finalizar as conexões com os bancos de dados de origem e destino.

```
private void closeDB() throws Exception {
    if (targetConn != null)
        targetConn.close();
    if (sourceConn != null)
        sourceConn.close();
}
```

Fonte: Autor.

O Quadro 15 apresenta o método startModel() que cria um modelo de banco de dados com as tabelas necessárias para a publicação e transferência dos dados de uma base de dados relacional para a base de dados RDF. A classe *ModelFactory*, integrante da biblioteca Jena, possibilita verificar se o modelo RDF existe ou não no banco de dados destino. Em caso negativo o mesmo é criado.

Quadro 15 - Método para criação de estrutura no banco de Dados.

```
private IDBConnection conn;  
private ModelMaker maker;  
private Model model;  
private String M_DB_MODEL;  
....  
private void startModel() throws Exception {  
    if (!conn.containsModel(M_DB_MODEL)) {  
        maker = ModelFactory.createModelRDBMaker(conn);  
        model = maker.createModel(M_DB_MODEL, true);  
    } else {  
        maker = ModelFactory.createModelRDBMaker(conn);  
        model = maker.openModel(M_DB_MODEL, true);  
    }  
}
```

Fonte: Autor.

Após a conexão do banco de dados e a verificação se o modelo RDF deve ou não ser criado torna-se necessária a migração dos dados. O desenvolvimento de rotinas de migração pode ser trabalhoso caso sejam consideradas as tabelas individualmente. Uma maneira de simplificar isto é recorrer ao metadados de um banco de dados relacional, permitindo que as tabelas e as colunas dessas tabelas sejam recuperadas. Deste modo é possível produzir rotinas genéricas de migração de informação. O Quadro 16 abaixo demonstra a função desenvolvida em Delphi para varrer a base de dados relacional, localizando a partir do metadados todas as tabelas. O nome das tabelas é então adicionado a uma lista que será utilizada em outras funções do protótipo.

Quadro 16 - Procedimento para localizar a meta informação no banco de dados de origem (Relacional).

```

procedure TForm1.Button1Click(Sender: TObject);
var
i,j: integer;
begin
  ADOQuery1.Close;
  ADOQuery1.SQL.Clear;

  ADOQuery1.SQL.Add(' SELECT c.relname AS Tabela ');
  ADOQuery1.SQL.Add(' FROM pg_class c ');
  ADOQuery1.SQL.Add(' LEFT JOIN pg_namespace n ON n.oid = c.renamespace ');
  ADOQuery1.SQL.Add(' LEFT JOIN pg_type t ON t.oid = c.reltypes ');
  ADOQuery1.SQL.Add('      where n.nspname = "public" ');
  ADOQuery1.SQL.Add('      and c.relkind = "r";');

  ADOQuery1.Open;
  ADOQuery1.First;
  ListBox1.Items.Clear;
  while not ADOQuery1.Eof do
  begin
    ListBox1.Items.Add(ADOQuery1.fieldbyname('Tabela').AsString);
    ADOQuery1.Next;
  end;
end;

```

Fonte: Autor.

No Quadro 17 pode ser verificada a função que realiza a inserção dos dados no modelo de tripla através do comando SQL de INSERT. Após a leitura de cada tabela constante na lista armazenada anteriormente recupera-se as colunas desta tabela que irão representar as propriedades no modelo RDF. Cada inserção é composta pelos valores que identificam o Sujeito, o Predicado e o Objeto. O Sujeito é em geral representado pelo identificador de uma tupla pertencente a uma tabela em particular. O Predicado é obtido através do nome da coluna de determinada tabela, enquanto que o Sujeito é obtido através do valor da coluna da tabela em questão.

De modo geral, com exceção da(s) coluna(s) que representa(m) a chave primária, o número de inserções no modelo RDF é definido pelo número de colunas de determinada tabela.

Quadro 17 - Função para realizar a Inserção no banco de dados RDF.

```

procedure TForm1.Button3Click(Sender: TObject);
begin
  Gauge.Progress:= 0;
  Gauge.MaxValue:= 0;
  ADOQuery1.Close;
  ADOQuery1.SQL.Clear;
  ADOQuery1.SQL.Add('select * from ' + ListBox1.Items[j]);
  ADOQuery1.Open;
  //Iniciando o Gauge
  Gauge.MaxValue:= ADOQuery1.RecordCount;
  while not ADOQuery1.Eof do
  begin
    for j := 0 to ADOQuery1.Fields.Count-1 do
    begin
      if j = 0 then
        chave := ADOQuery1.fields[j].AsString;
        comando := 'insert into jena_g1t1_stmt(subj, prop, obj, graphid) values(' +
          char(39) + prefixo + listbox1.Items[i] + '#' + chave + char(39) + ',' +
          char(39) + prefixo + listbox1.Items[i] + '#' +
            ADOQuery1.fields[j].fieldName + char(39) + ',' +
          char(39) + prefixo + listbox1.Items[i] + '#' +
            stringreplace(ADOQuery1.fields[j].AsString, '"', "", [])
            + char(39) + ',' + "1"');
        ListBox2.Items.Add(comando);
        Update;
        ADOQuery2.Close;
        ADOQuery2.SQL.Clear;
        ADOQuery2.SQL.Add(comando);
        ADOQuery2.ExecSQL;
      end;
    //Encerrando o Gauge depois dos Inserts
    Gauge.Progress:= Gauge.Progress + 1;
    ADOQuery1.Next;
  end;
end;
end;
end;

```

Fonte: Autor.

## 4.2 AVALIAÇÃO DOS DADOS DISPONIBILIZADOS

Após a conversão dos dados da base relacional para uma base de dados triplicadas foram efetuadas algumas consultas utilizando-se da linguagem SPARQL através do framework Jena®, visando avaliar a consistência do processo de triplicação. A seguir são apresentas as consultas com os respectivos resultados.

- a) A consulta constante no Quadro 18 extrai as Cidades com os nomes 'Criciúma', 'Gurupá', 'Irituia'. O resultado da seleção pode ser observado no Quadro 19.

Quadro 18 - Seleção de algumas cidades da base de triplas.

```
SELECT ?Codigo_Pais ?Abreviatura ?Estado ?Nome_Cidade
WHERE " +
{ ?x myp:abrev ?Abreviatura .
  ?x myp:i_estados ?Estado .
  ?x myp:nome ?Nome_Cidade .
  ?x myp:i_cidades ?Codigo_Cidade .
  FILTER ( ?Nome = myp:Criciuma || ?Nome = myp:Gurupa || ?Nome = myp:Irituia )
};
```

Fonte: Autor.

Quadro 19 - Resultado da seleção de Cidades.

Codigo_Cidade	Abreviatura	Estado	Nome_Cidade
myp:4204608	myp:CRI	myp:22	myp:Criciuma
myp:1503101	myp:GUR	myp:5	myp:Gurupa
myp:1503507	myp:IRI	myp:5	myp:Irituia

Fonte: Autor.

- b) Nesta seleção (Quadro 20), foram extraídos os dados dos estados, a partir das abreviaturas de seus nomes, sendo 'SC', 'RO' e 'AM'. O Quadro 21 apresenta o resultado da seleção.

Quadro 20 - Seleção de informações de alguns estados.

```
SELECT ?Codigo_Estado ?Abreviatura ?Numero_pais ?Nome_Estado
WHERE
{ ?x myp:abrev ?Abreviatura .
  ?x myp:nome ?Nome_Estado .
  ?x myp:i_paises ?Numero_pais .
  ?x myp:i_estados ?Codigo_Estado .
  FILTER ( ?Abreviatura = myp:SC || ?Abreviatura = myp:RO || ?Abreviatura = myp:AM )
}
```

Fonte: Autor.

Quadro 21 - Resultado da seleção das informações de estados.

Codigo_Estado	Abreviatura	Numero_pais	Nome_Estado
myp:22	myp:SC	myp:28	myp:Santa Catarina
myp:1	myp:RO	myp:28	myp:Rondonia
myp:3	myp:AM	myp:28	myp:Amazonas

Fonte: Autor.

- c) A seleção apresentada no Quadro 22 permite a obtenção das informações de um candidato em particular, neste caso, o candidato de nº 25777. O resultado pode ser visto no Quadro 23.

Quadro 22 - Seleção das informações do candidato 25777.

```

SELECT ?ID_Candidato ?Nome_candidato ?Sigla_partido ?Situacao
WHERE
{
  ?x myp:i_candidatos ?ID_Candidato .
  ?x myp:nome_urna ?Nome_candidato .
  ?x myp:sigla_partido ?Sigla_partido .
  ?x myp:situacao ?Situacao .
  FILTER ( ?ID_Candidato = myp:25777 )
}

```

Fonte: Autor.

Quadro 23 - Resultado da seleção do candidato 25777.

ID_Candidato	Nome_Urna	Sigla_partido	Situacao
myp:25777	myp:JOÃOZINHO	myp:DEM	myp:Indeferido

Fonte: Autor.

Até o momento foram realizadas consultas envolvendo somente uma unidade de informação, ou seja, o candidato ou país isoladamente. Para um melhor entendimento foram efetuadas duas consultas considerando mais de uma unidade de informação. O princípio básico neste tipo de operação sobre uma estrutura em grafo como o RDF envolve a utilização da variável capturada enquanto “Objeto”, em uma tripla subsequente como “Sujeito”. Deste modo, o Objeto passa para Sujeito criando uma ligação que permite a navegabilidade em um grafo, similar ao processo de junção em banco de dados relacional. A seguir são apresentas as consultas com os respectivos resultados:



- a) Na seleção constante no Quadro 24 apresenta a utilização de dois prefixos distintos identificando **Candidatos** e **Cidades**. Através do primeiro prefixo é possível obter os dados do candidato, inclusive o código na cidade que serve de referência, neste caso entendido como o Objeto da tripla. O segundo prefixo adicionado possibilita a ligação com as informações das cidades em que o Objeto da tripla anterior passa a ser o Sujeito da tripla seguinte com o intuito de recuperar o nome da cidade. O resultado obtido através do relacionamento entre candidato e cidade é apresentado no Quadro 25.

Quadro 24 - Seleção considerando informações de candidatos e cidades.

```
PREFIX myp: <http://DadosAbertos.org/candidatos#>
PREFIX myp2: <http://DadosAbertos.org/cidades#>
SELECT ?ID_Candidato ?Nome_candidato ?Sigla_partido ?Situacao ?Sexo
?Cidade
WHERE
{
  ?x myp:i_candidatos ?ID_Candidato .
  ?x myp:nome_urna ?Nome_candidato .
  ?x myp:sigla_partido ?Sigla_partido .
  ?x myp:situacao ?Situacao .
  ?x myp:i_cidades ?i_cidade .
  ?i_cidade myp2:nome ?Cidade .
  ?x myp:sexo ?Sexo
  FILTER ( ?Situacao = myp:Deferido && ?Sigla_partido = myp:PMDB)
}
```

Fonte: Autor.

Quadro 25 - Resultado da seleção envolvendo candidatos e cidades.

ID_Candidato	Nome_candidato	Sigla_partido	Situacao	Sexo	Cidade
myp:15	myp: ROMANA REMOR	myp:PMDB	myp:Deferido	myp:Feminino	myp2:Criciuma
myp:15015	myp: JOAQUIM FABRIS	myp:PMDB	myp:Deferido	myp:Masculino	myp2:Criciuma

Fonte: Autor.

- b) Na seleção demonstrada no Quadro 26 são utilizados dois prefixos distintos que representam **Cidades** e **Estados**. A partir do primeiro prefixo obtêm-se o nome da cidade e, considerando o segundo prefixo, recupera-se o nome do estado da cidade por meio do código do estado do primeiro prefixo. O resultado obtido é apresentado no Quadro 27.

### Quadro 26 - Seleção utilizando os prefixos de cidades e estados.

```
PREFIX myp: <http://DadosAbertos.org/cidades#>
PREFIX myp2: <http://DadosAbertos.org/estados#>
SELECT ?Codigo_Cidade ?Nome_Cidade ?Nome_estado
WHERE
{
  ?x myp:i_cidades ?Codigo_Cidade .
  ?x myp:nome ?Nome_Cidade .
  ?x myp:i_estados ?Sigla_partido .
  ?x myp:i_estados ?i_estado .
  ?i_estado myp2:nome ?Nome_estado .
}
```

Fonte: Autor.

### Quadro 27 - Resultado da seleção envolvendo cidades e estados.

Codigo_Cidade	Nome_Cidade	Nome_estado
myp:2207702	myp:Parnaíba	myp2:Piauí
myp:2300804	myp:Antônia do Norte	myp2:Ceará
myp:2918902	myp:Lajedão	myp2:Bahia

Fonte: Autor.

## 5 CONSIDERAÇÕES FINAIS

O objetivo geral deste trabalho foi propor um modelo de dados abertos a partir de uma fonte de informação governamental visando a democratização da informação e interoperabilidade entre sistemas computacionais.

A utilização deste modelo de disponibilização de dados permite que diversas organizações, não só de governo ou da área científica, mas também do meio corporativo, tornem públicos seus dados reforçando o que vem sendo caracterizado como uma tendência. Acredita-se que dentro de pouco tempo, a disponibilização de dados abertos seja o modelo padrão de fato no compartilhamento de informações na Web. A confirmação desta tendência promoveria a transformação de um espaço global de informação, para um espaço global de dados, formando a base para novas soluções e aplicações de conhecimento no futuro.

Dados governamentais ou de qualquer outra organização em formato aberto pode potencializar a inclusão digital, uma vez que promovem uma padronização facilitando assim o acesso por pessoas ou sistemas computacionais.

Este modelo de disponibilização tem o potencial de prover mais transparência uma vez que os dados podem ser reutilizados por terceiros, livremente permitindo uma visão muito mais ampla das ações do governo através da disponibilização de dados governamentais. A comparação de dados governamentais com dados de outras fontes pode conduzir a novas perspectivas na avaliação sobre o desempenho do governo e do impacto de suas ações.

Contudo, esse potencial é ainda pouco explorado, mas tende a evoluir a medida que os dados governamentais sejam liberados sem restrição de uso. Cada restrição adicionada aos dados abertos exclui a possibilidade de reutilização impactando diretamente na sua disponibilização e utilidade.

Outro benefício que dados abertos podem propiciar diz respeito à possibilidade de integração de dados de diferentes domínios, visto que disponibilizando dados usando padrões abertos facilita o compartilhamento de dados. A partir da integração entre diferentes fontes cria-se uma rede de dados potencializando descobertas úteis com o objetivo de melhoria da qualidade de vida da sociedade. Organizações do setor público ou privadas podem utilizar a informação dos mais variados órgãos públicos para prover valor agregado através da combinação de informação própria com outras fontes de dados.

Apesar dos dados abertos proporcionarem benefícios e do crescente interesse na disponibilização, ainda existem muitos problemas e desafios que precisam ser enfrentados para que isso se torne realidade. Antes de tudo, o sucesso dos dados abertos

governamentais requer uma mudança cultural, tanto por parte do governo quanto por parte da sociedade. Do lado governamental, é necessário que os gestores públicos entendam que os dados, em sua grande maioria, são de natureza pública e que, por este motivo, deveriam estar disponíveis para a sociedade como um todo.

Menciona-se ainda como desafio a questão da qualidade dos dados. O receio de disponibilizar dados de baixa qualidade tem levado a uma resistência por parte de alguns órgãos que temem a publicação de dados que são gerados a partir de sistemas de informação transacionais usados no dia a dia do governo.

De modo geral, é importante que se reconheça e perceba o valor e o potencial dos dados abertos em que o sucesso de uma iniciativa de dados abertos pode ser medido pela sua usabilidade a partir de sua publicação.

A partir da revisão de dados abertos, do ferramental e dos padrões, desenvolveu-se um protótipo capaz de transferir informações de um modelo tradicional utilizado para representar informação, neste caso o modelo relacional, para o modelo de dados abertos mais utilizado atualmente, representado através de triplas RDF. O protótipo possibilitou a validação do processo de publicação dos dados considerando o modelo de dados proposto para representar as informações de um domínio governamental, neste caso o Site DivulgaCand. Este site, disponibilizado pelo Tribunal Superior Eleitoral divulga informações sobre candidatos a eleições possibilitando um conjunto interessante de filtros para análise da informação. Contudo, a informação não se encontra em formato aberto recomendado pelo W3C, o que motivou a escolha desse site como cenário para a proposta e desenvolvimento do modelo apresentado neste trabalho.

Durante o desenvolvimento deste trabalho foram identificadas outras possibilidades para trabalhos futuros. Considerando o protótipo elaborado neste trabalho é possível identificar pontos de evolução no desenvolvimento de novas ferramentas e aplicações para a publicação de dados abertos, como exemplo, a flexibilização da interface, possibilidade de escolha das fontes de origem e destino, definição de prefixos para cada unidade de informação através da interface, entre outros.

Uma possibilidade identificada na revisão bibliográfica refere-se a ausência de conteúdo adequado que promova um estudo amplo das tecnologias de armazenamento de dados triplicado em formato RDF.

Como mencionado anteriormente, a publicação de dados em formato padronizado conduz a percepção de democratização e abre espaço para a criação de aplicações que possam melhorar serviços destinados aos cidadãos. Algumas aplicações foram pensadas como possibilidades para a extensão deste trabalho. Por exemplo, uma ferramenta que auxiliasse as pessoas que se utilizam de transporte

rodoviário. Nesta aplicação estariam disponíveis para fácil acesso todas as rotas das empresas de transporte coletivo, municipal, estadual ou nacional, com seus preços e taxas de embarque. Isto se justifica já que para pesquisar horários e itinerários é necessária a busca pelas informações em cada empresa individualmente e muitas vezes o usuário desconhece a empresa que atue em determinada rota.

## REFERÊNCIAS

ANTONIOU, G; HARMELEN, F. A Semantic Web Primer. 2. ed. **Massachusetts Institute of Technology**, 2008. 264 p.

APACHE. Apache Jena. Disponível em: [http://jena.apache.org/getting\\_started/index.html](http://jena.apache.org/getting_started/index.html). 2011. Acesso em: 20 out. 2013.

AUER, S.; BIZER, C.; KOBILAROV, G.; LEHMANN, J.; CYGANIAK, R.; IVES, Z. DBPedia: a nucleus for a web of open data. **Proceedings of the 6th International The Semantic Web and 2nd Asian Conference on Asian Semantic Web Conference**. p. 722-735, 2007.

BATISTA, A; SILVA, N; MIRANDA C. Infraestrutura Nacional De Dados Abertos, **Anais do VI Congresso de Gestão Pública**. 2013

BERGMAN, Mike. Advantages and Myths of RDF. Disponível em: <http://www.mkbergman.com/483/advantages-and-myths-of-rdf>. 2009. Acesso em: 25 ago. 2013.

BERNERS-LEE, T; HENDLER, J; LASSILA, O. The semantic web. **Scientific American**, v. 284, n. 5, p. 34-43, maio 2001.

BIZER, C; BERNERS-LEE, T; HEATH, T. Linked data: the story so far. **International Journal on Semantic Web and Information Systems**, v. 5, n. 2, p. 1-22, 2009.

BREITMAN, K. K. **Web semântica**: a internet do futuro. Rio de Janeiro: LTC, 2006.

BREITMAN, K. K.; CASANOVA, M. A.; TRUSZKOWSKI, W. **Semantic Web**: Concepts, Technologies and Applications. Springer, 2007

CAMATTA; GOBIRA; MELOTTI. Handbook de TI para Concursos: O guia Definitivo. 2011.

CARROLL, J. J.; DICKINSON, I.; DOLLIN, C.; REYNOLDS, D.; SEABORNE, A.; WILKINSON, K. Jena: Implementing The Semantic Web Recommendations. **Technical Report**, 2002.

DIETRICH, D.; GRAY, J.; MCNAMARA, T.; POIKOLA, A.; POLLOCK, R.; TAIT, J.; ZIJLSTRA, T. Open Data Handbook Documentation. 1<sup>st</sup> ed., 2012.

DINIZ, V. Como Conseguir dados Governamentais Abertos. **III Congresso de Gestão Pública**. 2010.

EAVES, D.. The Three Laws of Open Government Data, 2009 Disponível em: <<http://eaves.ca/2009/09/30/three-law-of-open-government-data/>>. Acesso em: 21 set. 2013.

EAVES, D Case Study: How Open data saved Canada \$3.2 Billion, 2010 Disponível em: <<http://eaves.ca/2010/04/14/case-study-open-data-and-the-public-purse/>> Acesso em : 12 dez. 2013

FENSEL, D.; HENDLER, J. A.; LIEBERMAN, H.; WAHLSTER, W. **Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential**, The MIT Press, 2005, 503 p.

FERREIRA, J.; SANTOS, P. O Modelo de Dados Resource Description Framework (RDF) e o seu Papel Na Descrição de Recursos. **Instituto de Informática Universidade Federal de Goiás**. v. 23, n. 2, p. 13-23, 2013.

FURGERI, S. O Papel das Linguagens de Marcação para a Ciência da Informação. **Transinformação**, v.18, p. 225-239. 2006.

HEATH, T; BIZER, C. Linked Data: Evolving the Web into a Global Data Space, 1 ed. New York, USA: Morgan & Claypool Publishers, 2011, 137p.

HOLDREN, J. P.; ORSZAG, P.; PROUTY, P. F.. Memorandum for Heads of Departments and Agencies. 2009. Disponível em: <[http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_fy\\_2009/m09-12.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_fy_2009/m09-12.pdf)>. Acesso em: 20 set. 2013.

IBM. Search RDF data with SPARQL. 2005. Disponível em: <<https://www.ibm.com/developerworks/xml/library/j-sparql/>>. Acesso em: 12 nov. 2013.

MAZONI, M. O papel dos Dados Abertos para seguir Construindo um Novo Brasil. **IV Consegi - Dados abertos Para a Democracia na Era Digital**, 2011.

MILLER, E. An Introduction to the Resource Description Framework. **D-Lib Magazine**, v. 4, n. 5, may 1998. Disponível em: <<http://www.dlib.org/dlib/may98/miller/05miller.html>>. Acesso em: 04 nov. 2013.

MILSTEAD, J; FELDMAN, S. Metadata: Cataloging by Any Other Name, 1999. Disponível em: <[http://www.iicm.tugraz.at/thesis/cguetl\\_diss/literatur/Kapitel06/Referenc es/Milstead\\_et\\_al.\\_1999/metadata.html](http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel06/Referenc es/Milstead_et_al._1999/metadata.html)>. Acesso em: 19 out. 2013.

NCE. URI, 2013. Disponível em: <<http://www.nce.ufrj.br/ginape/cursoshtml/conteudo/ligacoes/uri.htm>>. Acesso em: 22 nov. 2013.

OD4D. Gerenciamento do ciclo de abertura de Dados, 2013. Disponível em: <<http://www.od4d.org/2013/01/09/>>. Acesso em: 22 nov. 2013.

OPENGOVDATA. Open Government Data, 2007. Disponível em: <<http://www.opengovdata.org/>>. Acesso em: 25 jun. 2013.

RAMALHO, Rogério. Web Semântica: Aspectos Interdisciplinares da Gestão de Recursos Informacionais no Âmbito da Ciência da Informação. 2006. 121 f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília, 2006.

SAYÃO, L.; MARCONDES, C. O desafio da interoperabilidade e as Novas Perspectivas para as Bibliotecas Digitais. **Transinformação**, v. 20, n. 2, p. 133-148, 2008.

SHADBOLT, N.; HALL, W.; BERNERS-LEE, T. The Semantic Web Revisited. **IEEE Intelligent Systems**. v. 21, n.3, 2007.

SOUZA, R.; ALVARENGA, L. A Web Semântica e suas Contribuições Para a Ciência da Informação. **Ciência da Informação**, v.33, n.1, p.132-141, 2004.

TAUBERER, J. What is RDF. 2006. Disponível em: <<http://www.xml.com/pub/a/2001/01/24/rdf.html>>. Acesso em 06 Jun.2013



TSE. Divulgação de candidaturas disponível, 2012 em: <<http://www.tse.jus.br/eleicoes/eleicoes-2012/divulgacao-de-candidaturas-divulgacand2012>>. Acesso em: 05 Ago. 2013.

VICKERY, Grahah. Review of recent studies on PSI re-use and related market developments. Disponível em: [http://www.paikkatietoikkuna.fi/c/document\\_library/get\\_file?uuid=b1ad5545-266e-4e1b-8970-b855dbcbf997&groupId=108478](http://www.paikkatietoikkuna.fi/c/document_library/get_file?uuid=b1ad5545-266e-4e1b-8970-b855dbcbf997&groupId=108478). Acesso em: 20 Nov. 2012.

W3C. RDF Primer, 2004 Disponível em: <<http://www.w3.org/TR/rdf-primer/>>. Acesso em: 04 Nov. 2013.

W3C. SPARQL Query Language for RDF, 2008a. Disponível em: <<http://www.w3.org/TR/rdf-sparql-query/#grammar>>. Acesso em: 25 Mai. 2013.

W3C. Sobre o Consórcio W3C. 2008b. Disponível em: <<http://www.w3c.br/sobre/>>. Acesso em: 18 Mai. 2013.

W3. W3C Semantic Web Frequently Asked Questions, 2009. Disponível em: <<http://www.w3.org/2001/sw/SW-FAQ>>. Acesso em: 12 Jan. 2013.

W3C. Dados Abertos Governamentais. 2011. Disponível em: <<http://www.w3c.br/divulgacao/pdf/dados-abertos-governamentais.pdf>>. Acesso em: 20 jul. 2013.

W3C. Linked Data, 2013. Disponível em: <<http://www.w3.org/standards/semanticweb/data>>. Acesso em: 10 jan. 2013.

W3C. SPARQL 1.1 Query Language, 2013. Disponível: <<http://www.w3.org/TR/sparql11-query/>>. Acesso em: 12 Nov. 2013.

ZUIDERWIJK, A.; ANSSEN, M.; PARNIA, A. The complementarity of open data infrastructures: An analysis of functionalities. **Proceedings of the 14th Annual International Conference on Digital Government Research**. p. 166-171, 2013.